

NODE LINK SEARCHING DEVICE

Publication number: JP9218876 (A)

Publication date: 1997-08-19

Inventor(s): KUBO SHINYA

Applicant(s): NIPPON ELECTRIC CO

Classification:

- international: G06F12/00; G06F13/00; G06F17/30; G06F12/00; G06F13/00; G06F17/30; (IPC1-7): G06F17/30; G06F12/00

- European:

Application number: JP19960022343 19960208

Priority number(s): JP19960022343 19960208

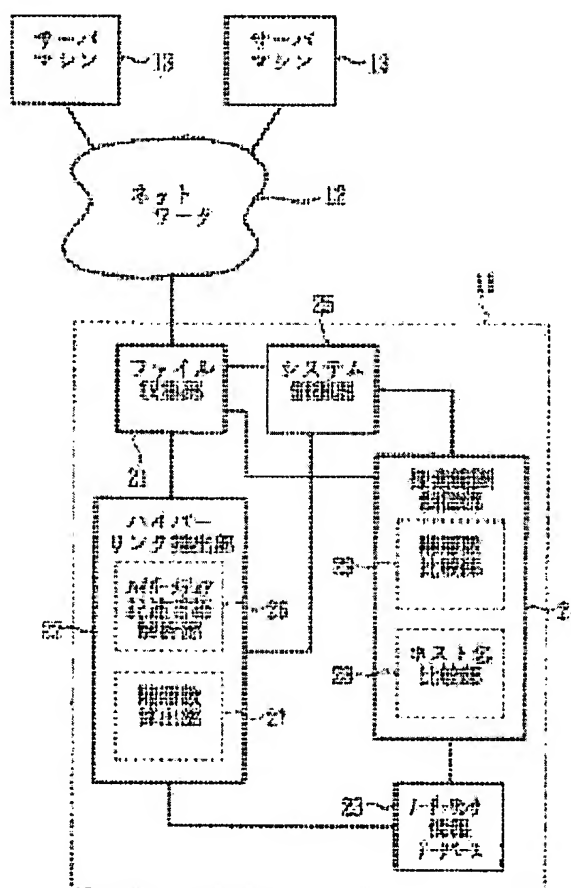
Also published as:

JP2940459 (B2)

Abstract of JP 9218876 (A)

PROBLEM TO BE SOLVED: To provide a node link searching device capable of limiting the search of unrequired nodes having no semantic connection.

SOLUTION: The number of the nodes from the node to be an origin to the node of a search destination is turned to a hierarchy number and the limiting conditions of a search range are set by the maximum value of the hierarchy number. A file collecting part 21 traces a link from the node to be the origin, successively reads the nodes from a server machine 13, obtains the title of the link destination node, a storage position and the correspondence relation of a link origin and a link destination from the nodes and registers them. A hierarchy calculation part 27 obtains the hierarchy number of the link destination node of the node every time the node is read from the server machine 13.; When the hierarchy number becomes more than the set maximum value, the tracing of the further link and the read of the nodes are stopped. Since the search range is limited by the hierarchy number from the origin, only the nodes with the semantic connection with the node of the origin are searched in a required range.



Data supplied from the esp@cenet database — Worldwide

(19)日本国特許庁 (J P)

(12) 公開特許公報 (A)

(11)特許出願公開番号

特開平9-218876

(43)公開日 平成9年(1997)8月19日

(51)Int.Cl. ⁸	識別記号	庁内整理番号	F I	技術表示箇所
G 0 6 F 17/30			G 0 6 F 15/40	3 1 0 C
12/00	5 4 5		12/00	5 4 5 A

審査請求 有 請求項の数4 O L (全 16 頁)

(21)出願番号 特願平8-22343

(22)出願日 平成8年(1996)2月8日

(71)出願人 000004237

日本電気株式会社

東京都港区芝五丁目7番1号

(72)発明者 久保 信也

東京都港区芝五丁目7番1号 日本電気株式会社内

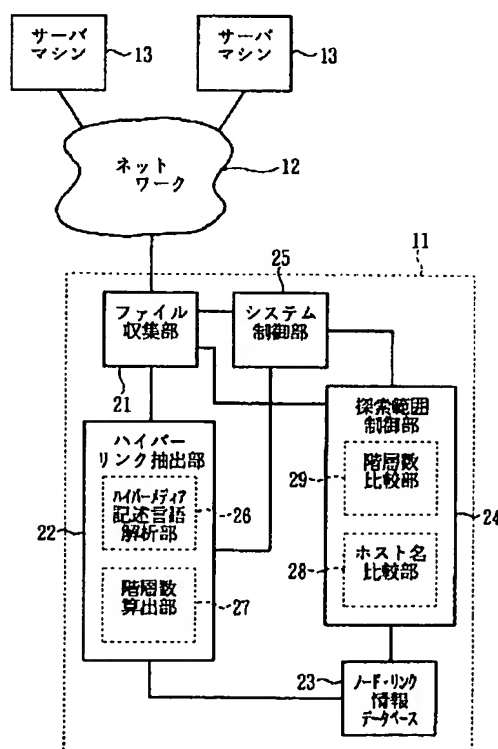
(74)代理人 弁理士 山内 梅雄

(54)【発明の名称】 ノード・リンク探索装置

(57)【要約】

【課題】 意味的なつながりのない不要なノードの探索を制限することのできるノード・リンク探索装置を提供する。

【解決手段】 起点となるノードから探索先のノードまでの間のノードの数を階層数とし、探索範囲の制限条件を階層数の最大値により設定する。ファイル収集部21は起点となるノードからリンクをたどりサーバマシン13より順次ノードを読み出し、このノードからそのリンク先ノードの名称や格納位置やリンク元とリンク先の対応関係を取得して登録する。階層算出部27はノードをサーバマシン13から読み込むごとにそのノードのリンク先ノードの階層数を求める。階層数が設定した最大値以上になったとき、それ以上先のリンクをたどりノードを読み出すことを中止する。起点からの階層数で探索範囲が制限されるので、起点のノードと意味的なつながりのあるノードのみを必要な範囲で探索できる。



【特許請求の範囲】

【請求項1】 ハイパーテキストの各ノードに含まれるリンク先のノードの名称とリンク先のノードの格納位置とを表わしたリンク情報を基にして任意のノードからリンク先のノードへの探索を順次行う際の探索範囲の制限条件を探索の起点となるノードから探索先のノードまでの間に存在するノードの数である階層数の最大値として設定する探索条件設定手段と、
前記リンク情報の示すリンク先のノードの内容をそれを蓄積しているサーバから読み出すことを前記探索の起点となるノードから順に繰り返す行うファイル収集手段と、
このファイル収集手段によって1つのノードの内容を読み込むごとにそのノードの含むリンク情報およびこれの示すリンク先のノードと今回読み込んだノードとの対応付けを表わす情報とを記憶するノード・リンク情報記憶手段と、
前記ファイル収集手段によって1つのノードの内容を読み込むごとにそのノードに含まれるリンク情報の示すリンク先のノードの階層数を求める階層数算出手段と、
この階層数算出手段によって求めた階層数が前記探索条件設定手段により設定した階層数の最大値よりも大きいとき今回読み込んだノード以降にリンクされているノードの内容の前記ファイル収集手段による読み込みを中止させる探索範囲制限手段とを具備することを特徴とするノード・リンク探索装置。

【請求項2】 ハイパーテキストの各ノードに含まれるリンク先のノードの名称とリンク先のノードの格納位置とを表わしたリンク情報を基にして任意のノードからリンク先のノードへの探索を順次行う際の探索範囲の制限条件を探索の起点となるノードから探索先のノードまでの間に存在するリンクの数である階層数の最大値として設定する探索条件設定手段と、
前記リンク情報の示すリンク先のノードの内容をそれを蓄積しているサーバから読み出すことを前記探索の起点となるノードから順に繰り返す行うファイル収集手段と、
このファイル収集手段によって1つのノードの内容を読み込むごとにそのノードの含むリンク情報およびこれの示すリンク先のノードと今回読み込んだノードとの対応付けを表わす情報とを記憶するノード・リンク情報記憶手段と、
前記ファイル収集手段によって1つのノードの内容を読み込むごとにそのノードに含まれるリンク情報の表わすリンク先のノードへのリンクの階層数を求める階層数算出手段と、
この階層数算出手段によって求めた階層数が前記探索条件設定手段により設定した階層数の最大値よりも大きいとき今回読み込んだノード以降にリンクされているノードの内容の前記ファイル収集手段による読み込みを中止

させる探索範囲制限手段とを具備することを特徴とするノード・リンク探索装置。

【請求項3】 探索の対象となるノードがネットワークに接続された複数のサーバに分散して格納されていることを特徴とする請求項1または請求項2記載のノード・リンク探索装置。

【請求項4】 前記ファイル収集手段は、今回読み込んだノードに含まれているリンク情報の示すリンク先のノードが既に読み込んだノードと同一であるか否かを判別する同一ノード判別手段と、この同一ノード判別手段により既に読み込んだノードと同一であると判別されたときそのノードの再度の読み込みを中止する多重読込中止手段とを具備することを特徴とする請求項1または請求項2記載のノード・リンク探索装置。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】本発明は、ノードとリンクからなるデータベースにおいて指定された条件の下で探索の対象とすべきノードの範囲を見出すノード・リンク探索装置に係わり、特に分散ハイパーメディアシステムのようにノードがネットワーク上の複数のサーバに分散して格納されたデータベースで用いられるノード・リンク探索装置に関する。

【0002】

【従来の技術】データベースの1つとして、ドキュメントファイルなど各種情報を格納した複数のノードをこれらの従属関係を表わしたリンクをたどることによって探索を行うハイパーテキストシステムがある。このようなハイパーテキストでの探索作業は、ノードに格納されている情報のインデックスと、ノード相互間のリンク関係を予め保持しておくことでスムーズに進めることができる。

【0003】特開平4-321144号公報には、ノード間相互間の関係を容易に把握できるように表示できるハイパーテキストシステムが開示されている。各ノードは任意のノードにリンクすることができるので、複数のノード間でループ状にリンクが形成されることがある。このシステムでは、ノード相互間のリンク関係を予め登録したテーブルを基にして表示画面を作成している。そして、ループ状になっているリンクを切り離し、あるノードを起点としたときに木構造としてノード相互間のリンク関係が表示されるようにしている。

【0004】従来、ハイパーテキストシステムはローカルのワークステーションなどに構築されていたが、近年の通信技術の発達により、ネットワークを介して接続された複数のサーバにノードを分散して格納するものが登場している。このようなシステムは、分散ハイパーメディアシステムと呼ばれている。たとえば、インターネット上の情報発信手段として、ワールド・ワイド・ウェブ（World Wide Web 以下WWWと表わす。）が注目され

ている。ハイパーメディアとは、文字や表などのテキストデータだけでなく、動画や音声などのマルチメディアデータも扱うことのできるハイパーテキストのことである。

【0005】分散ハイパーメディアシステムでは、各サーバに格納されているノードの情報をネットワークを介して取得することによって、ノード相互間のリンク関係を表わしたテーブルを作成するようになっている。ノード相互間のリンク関係を表わした情報をノード・リンク情報データベースと呼ぶことにする。また、ノード・リンク情報データベースを作成する装置をノード・リンク探索装置と呼ぶことにする。

【0006】ワールド・ワイド・ウェブ(WWW)では、ノードはインターネット上の複数のサーバマシンに分散して存在しており、各サーバに蓄積されているノードのデータはハイパーテキスト・トランスファ・プロトコル(HyperText Transfer Protocol 以下HTTPと表わす。)と呼ばれる手順に従って転送される。ノードになっているマルチメディアドキュメントは、ハイパーテキスト・マークアップ・ランゲージ(HyperText Markup Language 以下HTMLと表わす。)と呼ばれるハイパーメディア記述言語形式で記述されている。

【0007】ノード・リンク探索装置は、ネットワークを介して取得したHTMLで記述されているノードの内容を解析して、次のリンク先のノードの格納場所を表わすハイパーリンクを抽出する。ハイパーリンクは、HTMLで規定されているその開始位置を示す所定の文字列(これをタグと呼ぶ。)と終了位置を示すタグを検索することで、これらの間の文字列として抽出される。

【0008】リンク先ノードを表わすハイパーリンクは、ユニフォーム・リソース・ロケーターズ(Uniform Resource Locators 以下URLと表わす。)と呼ばれる表記形式により記述されている。取得したノードのテキスト中からハイパーリンクを示すタグに挟まれた部分を見出し、この中からURLで記述された文字列を抽出することによってノードとハイパーリンクの探索を行うことができる。ノード・リンク探索装置は、このような探索を繰り返すことによって、ノード・リンク情報データベースを構築する。ノード・リンク探索装置としては、“WWW Wanderer”、“WWW Robot”、“WWW Spider”などと呼ばれるものがある。

【0009】リンク先ノードを表わすハイパーリンクは、URLにおいて以下の構成が定義されている。
 <スキーム(Scheme)>: <スキーム特有部(Scheme-Specific-Part)>

ワールド・ワイド・ウェブ(WWW)でノードのドキュメントファイルの転送プロトコルとして用いられるハイパーテキスト・トランスファ・プロトコル(HTTP)スキームの場合には、ハイパーリンクは以下のように表される。

http://<host>;:<port>;:<path>;?<searchpart>;

【0010】ここで、“<host>;”はドキュメントファイルの存在するサーバマシンのホスト名を、“<port>;”は通信を行う際のポート番号を表わしている。ポート番号は省略可能であり、省略した場合には標準値“80”が指定されたものとして取り扱われる。“<path>;”はサーバマシン上のドキュメントファイルの存在する場所を表わしている。また、“<searchpart>;”はサーバマシンにデータを渡す場合に用いられる領域である。

【0011】図14は、ノードとこれらノード間を接続するハイパーリンクの一例を表わしたものである。ノード1001、ノード1002、ノード1003にそれぞれ含まれるドキュメントファイルは、HTMLで記述されている。ドキュメントファイルの内容は、各種のタグによりその内容が分類され識別可能になっている。このうちハイパーリンクを示すタグは、;...;で表され、“URL”に対応する部分がリンク先ノードを示している。ノード1001のドキュメントファイルでは、点線で囲んだ領域1004、1005がそれぞれハイパーリンクを表わしている。

【0012】ノード1001は、URLで“http://hostA/tale/TOC.html”のように表記されるノードであるとする。また、ノード1002は“http://hostB:80/tale/Introduction.html”と、ノード1003は“http://hostC:8080/tale/Birth.html”と表記されるノードであるものとする。ノード1001内のハイパーリンク1004は、ノード1002を指し示している。またノード1001内のハイパーリンク1005は、ノード1003を指し示している。このようにノードのドキュメントファイル内に登録されているハイパーリンクによって、リンク先のノードが表される。

【0013】図15は、ノードのドキュメント同士の関係の一例を表わしたものである。一点破線で囲んだ領域は、ノードの格納されているサーバを表わしている。領域1011はホスト名が“A”のサーバマシンを、領域1012はホスト名が“B”のサーバマシンを、領域1013はホスト名が“C”のサーバマシンをそれぞれ表わしている。

【0014】図14に示したノード1001～1003に含まれるドキュメントファイルのうち、HTMLで規定されたタグなどを取り除いた内容1014～1016はそのノードの格納されているサーバを表わす領域1011～1013内に表わしある。また、ノード1001のリンク先ノードがノード1002とノード1003であることをハイパーリンク1017、1018により表わしている。

【0015】次に、各ノードに登録されているハイパーリンクを基にしてノード・リンク情報データベースの構築を行うノード・リンク探索装置の構成を説明する。

【0016】図16は、従来から使用されているノード

・リンク探索装置の構成の概要を表わしたものである。ノード・リンク探索装置1021には、インターネットなどのネットワーク1022を介して複数のサーバマシン1023が接続されている。各サーバマシン1023には、ドキュメントファイルを格納したノードが蓄積されている。ノード・リンク探索装置1021は、サーバマシン1023からノードのファイルを収集するファイル収集部1031と、サーバマシン1023から取得したノードのファイルの中からハイパーリンクを抽出するハイパーリンク抽出部1032を備えている。

【0017】また抽出したハイパーリンクを基にしてノードの格納先などの各種属性情報およびノード相互間のリンク情報を蓄積・管理するノード・リンク情報データベース1033を有する。探索範囲制御部1034は、図示しない入力端末あるいは設定用の外部ファイルから与えられる探索範囲の条件に合致する範囲にノードの探索範囲を制限する部分である。システム制御部1035は、ノード・リンク探索装置1021内の各部の動作の流れを統括的に制御する回路部分である。

【0018】ハイパーリンク抽出部1032は、ハイパーメディア記述言語で書かれているノードのマルチメディアドキュメントファイルの内容を解析するハイパーメディア記述言語解析部1036を備えている。ハイパーメディア記述言語解析部1036は、各種タグを検出することによってハイパーリンクとリンク先ノードを抽出するようになっている。探索範囲は、サーバマシン1023の識別子としてのホスト名によって指定される。探索範囲制御部1034は、ノードの格納先のサーバマシンのホスト名と指定された探索範囲としてのホスト名とを比較するホスト名比較部1037を備えている。ノードの格納位置を表わす情報の一部としてのホスト名を比較することにより、ノードが探索範囲内のものであるか否かを判定するようになっている。

【0019】ノード・リンク探索装置1021のファイル収集部1031は、探索範囲として指定されたホスト名と一致するサーバマシン1023からネットワーク1022を通じてノードのマルチメディアドキュメントファイルを読み込む。そして、ハイパーリンク抽出部1032により読み込んだファイルからハイパーリンクとリンク先ノードの情報を抽出する。探索範囲制御部1034は抽出したリンク先ノードの格納されているサーバマシンのホスト名が探索範囲内の場合には、そのリンク先ノードに対して探索を継続する。一方、リンク先ノードの格納先のサーバマシンが探索範囲外のときは、それより先のノードへの探索は中止する。ノード・リンク情報データベース1033には、探索範囲内のノードおよびハイパーリンクについての各種属性が登録される。

【0020】このようにホスト名を探索範囲として指定するものの他に、何ら探索範囲を指定できないノード・リンク探索装置も存在する。

【0021】

【発明が解決しようとする課題】探索範囲を指定することのできないノード・リンク探索装置は、インターネットなどネットワーク全体を探索範囲とし、ネットワーク上の全てのサーバマシンに存在する全てのノードの探索を行う。このため、ノード・リンク情報データベースには、本来、探索範囲とすべきものの以外のノードに関する不必要な情報も蓄積され、その資源を浪費するばかりでなく、検索する際の作業効率も低下していた。さらに、ノードの存在するサーバマシンやネットワークをアクセスする時間が長くなり、他の利用者のネットワーク資源の利用が制限されてしまうという問題もある。

【0022】探索範囲を指定することのできる装置であっても、従来はホスト名を単位として探索範囲を制限することしかができない。これはノードの物理的な位置で探索範囲を制限していることになる。しかしながら、ハイパーテキストはノード間の意味的な関係に基づいて順次検索するものであるので、サーバマシンを単位とする探索範囲の指定では、適切な範囲指定ができない。

【0023】たとえば、図15に示した例では、サーバ“A”1011上に目次の登録されたノード1014が存在し、サーバ“B”1012上およびサーバ“C”1013上に、目次に対応する内容の文章の登録されたノード1015、1016がそれぞれ登録されている。このような場合、目次のみではなく、書かれている文書の中身までも探索範囲としたい場合には、サーバ“A”、サーバ“B”、サーバ“C”のすべてを探索範囲のホスト名として指定しなければならない。その結果、サーバ“B”、サーバ“C”上に存在する目次と何ら意味的なつながりの無い他の多数のノードまでが探索範囲となり、不要なノードの探索が行われてしまうという問題がある。

【0024】そこで本発明の目的は、意味的なつながりのない不要なノードの探索を制限することのできるノード・リンク探索装置を提供することにある。

【0025】

【課題を解決するための手段】請求項1記載の発明では、ハイパーテキストの各ノードに含まれるリンク先のノードの名称とリンク先のノードの格納位置とを表わしたリンク情報を基にして任意のノードからリンク先のノードへの探索を順次行う際の探索範囲の制限条件を探索の起点となるノードから探索先のノードまでの間に存在するノードの数である階層数の最大値として設定する探索条件設定手段と、リンク情報の示すリンク先のノードの内容をそれを蓄積しているサーバから読み出すことを探索の起点となるノードから順に繰り返す行うファイル収集手段と、このファイル収集手段によって1つのノードの内容を読み込むごとにそのノードの含むリンク情報およびこれの示すリンク先のノードと今回読み込んだノードとの対応付けを表わす情報とを記憶するノード・リ

リンク情報記憶手段と、ファイル収集手段によって1つのノードの内容を読み込むごとにそのノードに含まれるリンク情報の示すリンク先のノードの階層数を求める階層数算出手段と、この階層数算出手段によって求めた階層数が探索条件設定手段により設定した階層数の最大値よりも大きいとき今回読み込んだノード以降にリンクされているノードの内容のファイル収集手段による読み込みを中止させる探索範囲制限手段とをノード・リンク探索装置に具備させている。

【0026】すなわち請求項1記載の発明では、探索の起点となるノードから探索先のノードまでの間に存在するノードの数を階層数とし、探索を行う範囲を制限する条件を階層数の最大値により指定する。起点となるノードからリンクをたどることによって順次ノードの内容をサーバから読み出し、読み出したノードのリンク先のノードについてその名称や格納位置およびリンク元ノードとリンク先ノードとの対応関係を登録する。また、ノードをサーバから読み出すごとに読み込んだノードのリンク先のノードについての階層数を求める。この階層数が設定した最大値以上になったとき、それ以上先へリンクをたどりノードを読み出すことを中止する。

【0027】これらにより、起点としたノードからの階層数によって探索の範囲が制限される。このように起点のノードからの階層数によって探索範囲を制限しているので、探索されるノードは起点のノードと意味的なつながりの強いもののみとなる。また階層数により探索範囲を制限できるので、必要な範囲での情報のみを収集することができる。

【0028】請求項2記載の発明では、ハイパーテキストの各ノードに含まれるリンク先のノードの名称とリンク先のノードの格納位置とを表わしたリンク情報を基にして任意のノードからリンク先のノードへの探索を順次行う際の探索範囲の制限条件を探索の起点となるノードから探索先のノードまでの間に存在するリンクの数である階層数の最大値として設定する探索条件設定手段と、リンク情報の示すリンク先のノードの内容をそれを蓄積しているサーバから読み出すことを探索の起点となるノードから順に繰り返し行うファイル収集手段と、このファイル収集手段によって1つのノードの内容を読み込むごとにそのノードの含むリンク情報およびこれの示すリンク先のノードと今回読み込んだノードとの対応付けを表わす情報とを記憶するノード・リンク情報記憶手段と、ファイル収集手段によって1つのノードの内容を読み込むごとにそのノードに含まれるリンク情報の表わすリンク先のノードへのリンクの階層数を求める階層数算出手段と、この階層数算出手段によって求めた階層数が探索条件設定手段により設定した階層数の最大値よりも大きいとき今回読み込んだノード以降にリンクされているノードの内容のファイル収集手段による読み込みを中止させる探索範囲制限手段とをノード・リンク探索装置

に具備させている。

【0029】すなわち請求項2記載の発明では、探索の起点となるノードから探索先のノードまでの間に存在するリンクの数を階層数とし、探索を行う範囲を階層数の最大値により制限している。これにより起点のノードと意味的なつながりが強いノードについての情報だけを収集することができる。また階層数で指定された探索範囲内のノードについてだけその名称や格納位置ならびにノード相互間のリンク関係を収集できる。

【0030】請求項3記載の発明では、探索の対象となるノードがネットワークに接続された複数のサーバに分散して格納されている。

【0031】すなわち請求項3記載の発明では、探索の対象となる各ノードは、ネットワークを介して接続された複数のサーバに分散して格納されている。探索範囲を階層数によって制限しているので、ノードが複数のサーバに分散されていても、多数のサーバから必要なノードの情報だけを収集することができる。たとえば、サーバ単位でしか探索範囲を指定できない場合には、起点とするノードから必要な階層数を越えるノードのみならず、起点のノードとリンク関係のない無関係のノードまで収集される場合もある。階層数によって探索範囲を制限することにより、探索範囲のノードが複数のサーバに分散されていても、起点とするノードから意味的なつながりのある必要範囲のノードについての情報だけを収集することができる。

【0032】請求項4記載の発明では、ファイル収集手段は、今回読み込んだノードに含まれているリンク情報の示すリンク先のノードが既に読み込んだノードと同一であるか否かを判別する同一ノード判別手段と、この同一ノード判別手段により既に読み込んだノードと同一であると判別されたときそのノードの再度の読み込みを中止する多重読込中止手段とを具備している。

【0033】すなわち請求項4記載の発明では、1度読み込んだことのあるノードの再度の読み込みを防止している。これにより、ループした範囲を繰り返し探索することを回避することができる。

【0034】

【発明の実施の形態】

【0035】

【実施例】図1は、本発明の一実施例におけるノード・リンク探索装置の構成の概要を表わしたものである。ハイパーメディア構造上の探索手順は、探索を開始したノードを根とする木構造と見ることができる。そこで、根を始点として各ノードのある木の深さを階層数とし、探索範囲を始点からの階層数で制限するようになっている。

【0036】ノード・リンク探索装置11には、インターネットなどのネットワーク12を介してノードを蓄積した複数のサーバマシン13が接続されている。ノード

・リンク探索装置11は、サーバマシン13からノードのファイルを収集するファイル収集部21と、サーバマシン13から取得したノードのファイルの中からハイパーリンクとリンク先ノードを抽出するハイパーリンク抽出部22を備えている。

【0037】またノード・リンク探索装置11は、抽出したハイパーリンクとリンク先ノードの各種属性情報を蓄積・管理するノード・リンク情報データベース23を有する。探索範囲制御部24は、図示しない入力端末あるいは設定用のファイルから入力される探索範囲を指定するための条件に合致するようにノードの探索範囲を制限する部分である。システム制御部25は、ノード・リンク探索装置11内の各部の動作の流れを統括的に制御する回路部分である。

【0038】ハイパーリンク抽出部22は、ハイパーメディア記述言語で書かれたノードのマルチメディアドキュメントファイルの内容を解析するハイパーメディア記述言語解析部26を備えている。ハイパーメディア記述言語解析部26は、各種タグを検出することによってハイパーリンクとリンク先ノードを抽出するようになっている。また、ハイパーリンク抽出部22は、抽出したリンク先ノードの階層数を算出する階層数算出部27を備えている。階層数算出部27は、リンク元のノードの階層数に“1”を加えたものをリンク先ノードの階層数として求める。

【0039】探索範囲は、サーバマシン13の識別子としてのホスト名と、最大階層数により指定される。探索範囲制御部24は、ノードの格納先のホスト名と探索範囲として指定されたホスト名とを比較するホスト名比較部28と、そのノードの階層数と探索範囲の制限値としての最大階層数とを比較する階層数比較部29を備えている。探索範囲制御部24は、ノードの階層数、格納先のホスト名によって探索範囲を制限する。また、同一のノードの多重読み込みを防止するために、前回の検索からの経過時間により探索の範囲を制限することも行う。

【0040】ファイル収集部21は、ネットワーク12を通じてサーバマシン13からノードのマルチメディアドキュメントファイルを読み込む。ハイパーリンク抽出部22は、ハイパーメディア記述言語解析部26の解析結果を基にして、マルチメディアドキュメントファイルからハイパーリンクとリンク先ノードを抽出する。次に、階層数算出部27は、リンク元ノードに“1”を加えることにより、リンク先ノードまたはハイパーリンクの階層数を算出する。

【0041】探索範囲制御部24は、ノードあるいはハイパーリンクの階層数が指定された最大階層数以下か否か、あるいはホスト名が指定されたものと部分一致するか否かを基に探索範囲を制限する。探索範囲内のノードおよびハイパーリンクについては、それらの属性情報がノード・リンク情報データベース23に蓄積される。ノ

ード・リンク情報データベース23内でノードについての属性を登録するテーブルをノードテーブルと呼び、リンクに関する属性を登録するテーブルをリンクテーブルと呼ぶことにする。このように、階層数によって探索範囲を制限することで、不必要なノードについての属性情報の収集およびデータベースへの登録を防いでいる。

【0042】図2は、ノードテーブルの登録内容の一例を表わしたものである。ノードテーブル31には、図の左から、ノード識別子32、スキーム(Scheme)33、ホスト名34、ポート番号35、スキーム特有部(Scheme-Specific)36、階層数37、登録日時38、探索日時39、最終更新日時41が登録される。このうちノード識別子32は、図1では示していないデータベース・マネジメント・システム(DBMS)によって、ノードの属性情報をノードテーブルに登録する際に自動的に生成される識別子である。スキーム33、ホスト名34、ポート番号35およびスキーム特有部36は、ノードの格納位置を示すURLの文字列の各部分に対応するものである。階層数37は、検索開始位置からのノードの階層数を表わす項目である。

【0043】登録日時38は、ノードがデータベースに最初に登録された日時を示す。探索日時39は、このノードが探索された最新の日時を表わしている。最終更新日時41は、ノードのマルチメディアドキュメントファイルの最終更新日時を表わしている。図中、点線42で囲んだ探索日時39と、最終更新日時41の項目は、リンク元のノードとして登録されるときに更新される属性であり、点線43で囲んだ項目は、リンク元ノードのハイパーリンクにより指し示されているリンク先ノードとして登録されるときに更新される属性である。

【0044】図3は、リンクテーブルの登録内容の一例を表わしたものである。リンクテーブル51は、図の左から、リンク識別子52、リンク元ノード識別子53、リンク先ノード識別子54が登録される。これらの項目には、DBMSで自動的に割り当てられたノードの識別番号が登録される。リンク識別子52で示されるハイパーリンクの出所元のノードのノード識別子がリンク元ノード識別子53に、ハイパーリンクの指す先のノードのノード識別子がリンク先ノード識別子54として登録される。

【0045】ノード・リンク情報データベースのノードテーブル31には、探索の開始点となるノードの情報を予め少なくとも1つ登録しておく。この際、登録しておくべき項目は、ノード識別子32、スキーム33、ホスト名34、ポート番号35、スキーム特有部36、階層数37、登録日時38である。探索の起点となるノードの階層数には“0”を設定しておく。

【0046】図4は、ノード・リンク探索装置の行う動作の流れを表わしたものである。一点破線61で囲んだステップの処理は、ノード・リンク装置のシステム制御

部25によって行われる。一点破線62で囲んだステップの処理は探索範囲制御部24により、一点破線63で囲んだステップの処理はファイル収集部21およびハイパーリンク抽出部22によって行われる。まず、探索に先立ちシステム制御部は探索条件の入力などのシステムの初期化を行う(ステップS101)。探索条件は、図1で示していない入力端末や外部ファイルから入力される。探索条件として、ここでは探索範囲とすべきノードの階層数の最大値(最大階層数)と、前回の探索結果を古いものと判定するための経過時間と、探索対象とするサーバマシンを特定するための文字列(探索するホスト名の一部あるいは全部)が条件として設定される。

【0047】システム制御部25は、ノードテーブル31を検索して、未探索のノードが存在するか否かを調べる(ステップS102)。ノードテーブル31中に登録されている探索日時39が、未定義(未登録)の場合、探索日時39が現在時刻よりも探索条件として設定された経過時間以上過去の時刻である場合のいずれかに該当するとき未探索のノードと判定する。未探索のノードが存在する場合には(ステップS102; Y)、探索範囲制御部24は、探索されていない1つのノードの属性情報をノードテーブル31から取り出す。この際、階層数の小さいものから優先的にノードを選択する(ステップS103)。したがって、最初は、階層数が“0”である探索開始点のノードの属性情報が取り出される。

【0048】次に、選択したノードが指定された探索範囲内に存在するノードであるかどうかを判別する(ステップS104)。探索範囲内か否かを判別する処理については後に詳細に説明する。探索範囲内に存在するノードである場合には(ステップS104; Y)、そのノードのドキュメントファイルをネットワークを通じて読み出し、これに記述されているリンク先ノードをノードテーブルに追加登録する等のテーブル情報更新処理(ステップS105)を行う。テーブル更新処理の詳細については後に説明する。1つのノードについての更新処理(ステップS105)を終えた後、再びステップS102に戻り、未探索のノードについての探索処理を繰り返す。

【0049】選択したノードが指定された探索範囲外のノードである場合には(ステップS104; N)、ノードテーブル31内の選択したノードについての探索時刻39を現在時刻に変更し(ステップS106)、ステップS102に戻る。未探索のノードがノードテーブルに存在しなくなったとき(ステップS102; N)、探索を終了するための終了処理を行い(ステップS107)、処理を終了する(エンド)。

【0050】図5は、探索範囲内に存在するノードであるか否かを判定する際の処理の流れを表わしたものである。ここでは、階層数のみによって探索範囲内か否かを判別している。まず探索範囲制御部24は、選択したノ

ードの階層数と探索条件として設定されている最大階層数とを比較する(ステップS201)。ノードの階層数が最大階層数よりも大きい場合には(ステップS201; N)、このノードは探索範囲に存在しないものと判定する(ステップS202)。ノードの階層数が最大階層数より小さいかあるいは等しい場合には(ステップS201; Y)、このノードが探索範囲に存在するものと判定する(ステップS203)。

【0051】図6は、探索範囲内に存在するノードであるか否かを判定する処理の他の一例の流れを表わしたものである。ここでは、階層数による判別の他に、ホスト名による判別を加えている。探索範囲制御部24は、選択したノードの階層数と最大階層数とを比較し(ステップS301)、ノードの階層数が最大階層数よりも大きい場合には(ステップS301; N)、このノードは探索範囲に存在しないものと判定する(ステップS302)。ノードの階層数が最大階層数より小さいかあるいは等しい場合には(ステップS301; Y)、探索範囲のホスト名として指定された文字列と、ノードのホスト名とが部分一致するか否かを調べる(ステップS303)。

【0052】たとえば、探索文字列として“AB”が指定されたときには、“ABC”や“ABD”などホスト名の先頭から指定された文字列“AB”を含むものはすべて部分一致していると判別される。“CAB”などのように部分一致していない場合には(ステップS303; N)、このノードを探索範囲に存在しないものと判定する(ステップS302)。部分一致する場合には(ステップS303; Y)、このノードが探索範囲に存在するものと判定する(ステップS304)。ここでは、階層数の判定を行ってからホスト名の部分一致を判定したが、これらの順序を入れ換えて行ってもよい。

【0053】図7は、図4に示したテーブル更新処理の流れを表わしたものである。この処理は、ファイル収集部21とハイパーリンク抽出部22により行われる。まず、図4のステップS104により探索範囲内に存在すると判別されたノードの内容をネットワークを通じてそれが格納されているサーバマシンから読み出すことを行う。このため、ファイル収集部21は、該当するノードの属性情報をノード・リンク情報データベース23のノードテーブル31から読み込む(ステップS401)。読み込んだノードの階層数をここでは、仮にnとする。次に、ファイル収集部21は、サーバマシンとの通信に用いる図示しない通信用バッファと一時バッファを初期化した後サーバマシン13との接続を行い、ノードのドキュメントファイルの転送(読み込み)準備を行う(ステップS402)。

【0054】続いてファイル収集部21は、分散ハイパーメディアシステムの転送プロトコルに従って、ドキュメントファイルの転送を開始する。ファイル収集部21

は、転送プロトコルメッセージヘッダ部を読み込み（ステップS403）、ドキュメントファイルの記述言語などのフォーマットを調べる（ステップS404）。読み込んだファイルがハイパーメディア記述言語形式でない場合には（ステップS404；N）、そのドキュメントファイルの終端まで読み込む（ステップS405）。そして、通信用バッファおよび一時バッファの解放ならびにサーバマシンとの接続を断するなどの後処理を行い（ステップS406）、処理を終了する（エンド）。

【0055】読み込んだファイルがハイパーメディア記述言語形式の場合には（ステップS404；Y）、ファイル収集部21は、転送プロトコルメッセージ本体のドキュメントファイルを通信用バッファに読み込む（ステップS407）。次に、ハイパーメディア記述言語解析部26により、ハイパーリンクの部分を表わすタグの開始文字から終了文字までを一時バッファに移動させる（ステップS408）。一時バッファに格納したハイパーリンクの中からリンク先ノードを記述している部分を取り出す（ステップS409）。階層数算出部は、リンク元のノードの階層数“n”に“1”を加えた“n+1”を、リンク先ノードの階層数として求める（ステップS410）。

【0056】このようにして得たノードとハイパーリンクの属性情報を、ノード・リンク情報データベース23内のノードテーブル31およびリンクテーブル51に書き込む（ステップS411）。この際、リンク元ノードの属性情報、リンク先ノードの属性情報の各一部項目と、ハイパーリンクの属性情報（リンクテーブル）の全項目を書き込む。これにより、今回読み込んだノードにリンクされている1つ階層の進んだリンク先ノードについての属性情報と、読み込んだノードからそのリンク先ノードへのハイパーリンクを各テーブルに1つ追加登録したことになる。ただし、同一のノードについて既に登録されている場合には、テーブルへの追加登録は行わず、探索日時などの更新のみを行う。これにより同一のノードやハイパーリンクの多重登録が回避される。

【0057】次に、今回読み込んだノードのドキュメントファイルの終端まで処理を行ったか否かを調べ（ステップS412）、終端に到らないときは（ステップS412；N）、ステップS407に戻る。各ノードは、複数のリンク先ノードを有することがあるので、ファイルの終端までこのような処理を繰り返し行うことにより、今回読み込んだノードのリンク先ノードの全てについてノードテーブルとリンクテーブルへの登録を行う。ファイルの終端まで処理したときは（ステップS412；Y）、このノードについてのテーブル更新処理を終了する（エンド）。

【0058】ここで、図7のステップS411においてノードテーブルとリンクテーブルに登録する属性情報の内容について説明する。サーバマシンから読み込んだノ

ードをリンク元ノードとし、このノードに含まれるハイパーリンクに記述されているノードをリンク先ノードとする。まず、リンク元ノードについては、探索日時39と、最終更新日時41を登録する。これにより、ノードをいつ検索したかの最新の時刻情報を残すことができる。たとえば、図2に示すノードテーブル31において、ノード識別子が“1”のノードをサーバマシンから読み出し、このハイパーリンクの指す先としてノード識別子が“2”のノードが記述されていたものとする。この場合は、ノード識別子が“1”のノードがリンク元ノードであり、ノード識別子が“2”のノードがリンク先ノードである。

【0059】図2において点線42で囲まれている探索日時と最終更新日時がリンク元ノードの属性情報として更新される。リンク先ノードの属性情報として登録されるのは図2の点線44で囲まれている部分である。すなわち、ノード識別子と、ノードの格納位置を示すURLの文字列に対応した、スキーム、ホスト名、ポート番号、スキーム特有部、階層数算出部で求めた階層数、およびデータベースに登録された日時である登録日時である。ノード識別子は、DBMSによって自動生成されたものが登録される。リンク先のノードについては、まだ実際にサーバマシン13から読み出してそれに含まれるハイパーリンクを調べていないので、探索日時および最終更新日時は未登録のままとなる。

【0060】リンクテーブル51には、DBMSによって自動生成されたリンク識別子52と、リンク元ノード識別子53と、リンク先ノード識別子54が登録される。図3を例に説明する。ノード識別子が“1”のノードを読み込み、このハイパーリンクによってノード識別子が“2”のノードがリンク先ノードとなっているものとする。まず、リンク識別子が“1”のハイパーリンク（55）の出所元のノードのノード識別子は“1”であるので、この値をリンク元ノード識別子（56）として登録する。またハイパーリンクの指す先のノードのノード識別子は“2”であるので、この値をリンク先ノード識別子（57）として登録する。読み込んだノードの他のハイパーリンクが登録されている場合には、それらリンクについてもリンク識別子を割り当て、リンク元ノード識別子とリンク先ノード識別子が登録される。2つハイパーリンクが存在している場合には、図3の点線58で示した範囲の情報がリンクテーブル51に登録される。

【0061】ノードテーブル31およびリンクテーブル51の更新は、1つのノードをサーバマシン13から読み込み、新たなリンク先を見出すたびに行われる。したがって、図4の流れ図において1つのノードについてテーブル更新処理（ステップS105）を終えた後、ステップS102に戻ると、テーブル更新処理において新たに登録されたノードもその探索対象になる。これによ

り、次々とリンク先への探索が進められる。ただし、その探索範囲は階層数などの条件によって制限される。

【0062】また、探索日時が未定義か経過時間以上過去の時刻であることを未探索判定基準にしており、また一度探索したノードの探索日時には現在時刻に近い時刻が登録されるので、同一のノードについて重ねてリンク先の調査が行われることはない。これにより、たとえばノード“A”のリンク先がノード“B”で、ノード“B”のリンク先がノード“A”のようにリンクによりループが形成されている場合であっても、ノード“A”を再度調べることが無く、最大階層数までループを繰り返したようなことがない。

【0063】このようにノードの階層数によって探索の範囲を制限しているため、起点となるノードと意味的なつながりの強いノードだけを探索することができる。また、サーバ単位でしか探索範囲を制限できない場合に比べて不必要な探索を低減することができる。

【0064】変形例

【0065】これまで説明した実施例では、ノードの属性情報として階層数を持たせているが、変形例ではハイパーリンクの階層数を基にして探索範囲内か否かを判別するようになっている。装置の構成は図1に示したものと同一でありその説明を省略する。

【0066】図8は、変形例のノード・リンク探索装置で用いられるノードテーブルの登録内容の一例を表わしたものである。図2と同一の項目には同一の符号を付してあり、それらの説明を適宜省略する。ノードテーブル71は、図2に示したノードテーブル31に比べて、階層数と探索日時を登録する項目が削除されている点で相違する。

【0067】図9は、変形例のノード・リンク探索装置で用いるリンクテーブルの登録内容の一例を表わしたものである。図3と同一項目には同一の符号を付してあり、それらの説明を適宜省略する。リンクテーブル81は、図3に示したリンクテーブル51に加えて、階層数の項目82と、探索日時の項目83を備えている。これらは、実施例においてはノードテーブルに登録されていたものである。変形例では、探索に先立って、探索の開始点となるハイパーリンクを少なくとも1つリンクテーブル81に登録しておく必要がある。ここで登録される探索開始点となるハイパーリンクは、仮想的なものであり、リンク元ノードが存在しない（不明）。また、探索の開始点となるハイパーリンクの示すリンク先ノードについての属性情報をノードテーブル71に予め登録しておかなければならない。

【0068】探索に先立ってノードテーブル71に登録する内容として、まず探索の起点のハイパーリンクの示すリンク先ノードのノード識別子がある。この値はDBMSによって自動的に割り当てられる。さらに、ノードの格納位置を示すURLの文字列に対応した項目とし

て、スキーム、ホスト名、ポート番号、スキーム特有部を初期登録しておく。また、このノードがデータベースに登録された時刻を表わす登録日時を初期登録する。図8の例では、点線72で囲んだ範囲の項目が初期登録される。

【0069】リンクテーブル81に初期登録しておく内容としては、探索の起点となるハイパーリンクの識別子であるリンク識別子がある。この値は、DBMSによって自動生成される。探索の起点となるハイパーリンクのリンク元は不明であるため、リンク元ノード識別子の初期値は未定義を表わす“0”とする。リンク先ノード識別子は、リンク先となるノードに対してDBMSの割り当てたノード識別子と同一の値を登録しておく。階層数は、探索の開始点である“0”を初期登録する。また探索日時は、当該ハイパーリンクについての探索を行った日時を登録するものであり、探索開始の初期値としては未定義のままとする。図9の例では点線84で囲んだ項目が初期登録される。

【0070】図10は、変形例におけるノード・リンク探索装置の行う処理の流れを表わしたものである。一点破線101で囲まれたステップは、システム制御部25の行う処理を表わしている。一点破線102で囲まれたステップは、探索範囲制御部24により、一点破線103で囲まれたステップは、ファイル収集部21およびハイパーリンク抽出部22によって行われる処理を表わしている。

【0071】まず、システム制御部25は、当該システムの初期化を行う（ステップS501）。この際、探索範囲とするハイパーリンクの階層数の最大値としての最大階層数と、前回の探索結果を古いものとして扱う基準となる経過時間とを設定する。さらにホスト名によってサーバの範囲を設定する場合には、ホスト名を制限するための文字列を入力する。これらは、図示しない入力端末あるいは外部ファイルから取り込む。次に、リンクテーブル81の中に探索していないハイパーリンクが存在するか否かを調べる（ステップS502）。探索されていないハイパーリンクとは、探索日時が未定義（未登録）のもの、あるいは探索日時が現在時刻よりも経過時間以上古いものである。

【0072】探索されていないハイパーリンクがリンクテーブル81に存在するときは（ステップS502；Y）、探索されていないハイパーリンクのうちの1つを選択しその属性情報を取り出す（ステップS503）。次に、探索範囲制御部24は、選択した1つのハイパーリンクが探索範囲内であるかどうかを判定する（ステップS504）。階層数あるいは階層数とホスト名の双方により判定されるが、その詳細については後に説明する。

【0073】探索範囲内に存在する場合には（ステップS504；Y）、テーブル更新処理（ステップS50

5)を行う。この処理では、ハイパーリンクの示すリンク先ノードの属性情報をノードテーブル71から取り出し、このノードのドキュメントファイルをサーバマシンから読み出し、これに登録されたリンク先ノードの属性情報を追加登録する等を行う。処理の詳細な流れについては後述する。ハイパーリンクの示すリンク先ノードについてのテーブル更新処理(ステップS505)を終えた後、再びステップS502に戻り、未探索のハイパーリンクについての探索処理を繰り返す。

【0074】選択したハイパーリンクが探索範囲外の場合には(ステップS504;N)、リンクテーブル81中の選択したハイパーリンクについての探索日時39を現在時刻に変更し(ステップS506)、ステップS502に戻る。未探索のハイパーリンクがリンクテーブル81に存在しなくなったとき(ステップS502;N)、探索を終了するための終了処理を行い(ステップS507)、処理を終了する(エンド)。

【0075】図11は、ハイパーリンクが探索範囲内か否かを判定する際の処理の流れを表わしたものである。ここでは、ハイパーリンクの階層数のみによって探索範囲内か否かを判別している。まず、探索範囲制御部24は、選択したハイパーリンクの階層数と探索条件として設定されている最大階層数とを比較する(ステップS601)。ハイパーリンクの階層数が最大階層数よりも大きい場合には(ステップS601;N)、このハイパーリンクは探索範囲に存在しないものと判定する(ステップS602)。ハイパーリンクの階層数が最大階層数より小さいかあるいは等しい場合には(ステップS601;Y)、このハイパーリンクが探索範囲に存在するものと判定する(ステップS603)。

【0076】図12は、ハイパーリンクが探索範囲内か否かを判定する処理の他の一例の流れを表わしたものである。ここでは、ハイパーリンクの階層数による判別の他に、ホスト名による判別を加えている。探索範囲制御部24は、選択したハイパーリンクの階層数と最大階層数とを比較し(ステップS701)、ハイパーリンクの階層数が最大階層数よりも大きい場合には(ステップS701;N)、このハイパーリンクは探索範囲に存在しないものと判定する(ステップS702)。ハイパーリンクの階層数が最大階層数より小さいかあるいは等しい場合には(ステップS701;Y)、このハイパーリンクの示すリンク先ノードについての属性情報をノードテーブル71から取り出す(ステップS703)。

【0077】取り出したリンク先ノードの属性情報に含まれるホスト名が探索範囲のホスト名として指定された文字列と、部分一致するか否かを調べる(ステップS704)。部分一致する場合には(ステップS704;Y)、先のハイパーリンクが探索範囲内に存在するものと判定する(ステップS705)。部分一致しない場合には(ステップS704;N)、ハイパーリンクが探索

範囲内に存在しないと判定する(ステップS702)。ここでは、階層数の判定を行ってからホスト名の部分一致を判定したが、これらの順序を入れ換えて行ってもよい。

【0078】図13は、図10に示したテーブル更新処理の流れを表わしたものである。この処理は、ファイル収集部21とハイパーリンク抽出部22により行われる。まず、図10のステップS504により探索範囲内に存在すると判別されたハイパーリンクの属性情報をリンクテーブル81から読み込む(ステップS801)。読み込んだハイパーリンクの階層数をここでは、仮にnとする。次に、このハイパーリンクのリンク先ノード識別子に対応するノードの属性情報をノードテーブル71から読み込む(ステップS802)。

【0079】ファイル収集部21は、サーバマシンとの通信に用いる図示しない通信用バッファと一時バッファを初期化した後サーバマシンとの接続を行い、ハイパーリンクの示すリンク先ノードのドキュメントファイルの読み込み準備を行う(ステップS803)。続いてファイル収集部21は、分散ハイパーメディアシステムの転送プロトコルに従って、ドキュメントファイルの転送を開始する。

【0080】ファイル収集部21は、転送プロトコルメッセージヘッダ部を読み込み(ステップS804)、ドキュメントファイルの記述言語がハイパーメディア記述言語形式でない場合には(ステップS805;N)、そのドキュメントファイルの終端まで読み込む(ステップS806)。そして、通信用バッファおよび一時バッファの解放ならびにサーバマシンとの接続を断るなどの後処理を行い(ステップS807)、処理を終了する(エンド)。

【0081】読み込んだファイルがハイパーメディア記述言語形式の場合には(ステップS805;Y)、ファイル収集部21は、転送プロトコルメッセージ本体のドキュメントファイルを通信用バッファに読み込む(ステップS808)。次に、ハイパーリンクの部分を表わすタグの開始文字から終了文字までを一時バッファに移動させ(ステップS809)、一時バッファに格納したハイパーリンクの中からリンク先ノードを記述している部分を取り出す(ステップS810)。階層数算出部27は、抽出したハイパーリンクの階層数として“n+1”を設定する(ステップS812)。

【0082】このようにして得たノードとハイパーリンクの属性情報を、ノード・リンク情報データベース23内のノードテーブル71およびリンクテーブル81に書き込む(ステップS812)。この際、リンク元ノードの属性情報、リンク先ノードの属性情報の各一部項目と、ハイパーリンクの属性情報の全項目を書き込む。これにより、今回の読み込んだノードのリンク先ノードについての属性情報と、読み込んだノードからそのリンク

先ノードへのハイパーリンクについて登録が行われる。ただし、同一のリンク元とリンク先を有するハイパーリンクが既に登録されている場合には、テーブルへの追加登録は行わず、探索日時などの更新のみを行う。これにより同一のノードやハイパーリンクが多重登録されることが回避される。

【0083】次に、今回読み込んだノードのドキュメントファイルの終端まで処理を行ったか否かを調べ（ステップS813）、終端に到らないときは（ステップS813；N）、ステップS808に戻る。各ノードは、複数のリンク先ノードを有することがあるので、ファイルの終端までこのような処理を繰り返し行うことにより、今回読み込んだノードのリンク先ノードの全てについてノードテーブルとリンクテーブルへの登録が行われる。ファイルの終端まで処理したとき（ステップS813；Y）、処理を終了する（エンド）。

【0084】ここで、図13のステップS812においてノードテーブル71とリンクテーブル81に登録される属性の内容について説明する。サーバマシンから読み込んだノードをリンク元ノードとし、このノードに含まれるハイパーリンクに記述されているノードをリンク先ノードとする。まず、リンク元ノードについては、最終更新日時41を登録する。たとえば、図8に示すノードテーブル71において、ノード識別子が“1”のノードをサーバマシンから読み出し、このハイパーリンクの指す先としてノード識別子が“2”のノードが記述されていたものとする。この場合は、ノード識別子が“1”のノードがリンク元ノードであり、ノード識別子が“2”のノードがリンク先ノードになる。

【0085】図8において点線73で囲まれている最終更新日時がリンク元ノードの属性情報として更新される。リンク先ノードの属性情報として登録されるのは図8の点線74で囲まれている部分である。すなわち、ノード識別子と、ノードの位置を示すURLの文字列に対応した、スキーム、ホスト名、ポート番号、スキーム特有部、およびデータベースに登録された日時である登録日時である。ノード識別子は、DBMSによって自動生成されたものが登録される。

【0086】リンクテーブル81には、DBMSによって自動生成されたリンク識別子と、リンク元ノード識別子と、リンク先ノード識別子と、階層数と、探索日時が登録される。図9を例に説明する。リンク識別子が“1”のハイパーリンクを基にしてリンク先であるノード識別子が“1”のノードを読み込んだものとする。読み込んだノードに記述されているハイパーリンクによりノード識別子が“2”とノード識別子が“3”のノードとがリンクされているものとする。

【0087】この際、ハイパーリンク84を参照して探索を行ったので、その時刻を探索日時85として登録する。次に新たに得られたハイパーリンクについてのリン

ク識別子をDBMSから取得し、これを追加登録するリンク識別子の欄に登録する。たとえば、リンク識別子が“2”のハイパーリンク（86）のリンク識別子（87）として“2”を登録する。この値は、テーブルの登録順などを基準にDBMSにより適宜与えられるID番号である。

【0088】ハイパーリンク86のリンク元ノードは、そのノード識別子の値が“1”であるので、リンク元ノード識別子88として“1”を登録する。またリンク先ノード識別子89には、リンク先のノードの属性をノードテーブル71に登録する際にDBMSにより与えられたノード識別子の値、すなわち“2”を登録する。また、階層数91には、階層数“0”のハイパーリンクをたどって得たノードから取得したハイパーリンクであるので“0”に“1”を加えた値“1”を登録する。ノード識別子が“3”のノードに向けてのハイパーリンク92についても同様の手順により登録される。追加登録したハイパーリンクについて図9の点線93で示した範囲の情報が登録されることになる。

【0089】ノードテーブル71およびリンクテーブル81の更新は、1つのハイパーリンクを基にしてそのリンク先ノードをサーバマシンから読み込み、新たなリンク先を見出すたびに行われる。したがって、図10の流れ図において1つのノードについてテーブル更新処理（ステップS505）を終えた後、ステップS502に戻ると、テーブル更新処理において新たに登録されたハイパーリンクもその探索対象になる。これにより、次々とリンク先への探索が進められる。ただし、その探索範囲は階層数などの条件によって制限される。

【0090】また、探索日時が未定義か経過時間以上過去の時刻であることを未探索の判定基準にしているのので、一度探索したハイパーリンクについて再度調べることがない。これにより、たとえばノード“A”のリンク先がノード“B”で、ノード“B”のリンク先がノード“A”のようにリンクによりループが形成されている場合であっても、ノード“A”からノード“B”へのハイパーリンクを再度調べることが無く、最大階層数までループをたどるようなことがない。

【0091】このようにハイパーリンクの階層数を基に探索範囲を制限しても、ノードの階層数を基に探索範囲を制限したときと同様の効果を得ることができる。ただし、ハイパーリンクに階層数を付与する場合には、リンクテーブルを参照して得たハイパーリンクのリンク先ノードを調べるために、ノードテーブルも参照する必要がある。実施例のようにノードに階層数を付与した場合に比べてテーブルの参照処理が増加する。

【0092】

【発明の効果】以上詳細に説明したように請求項1ないし請求項3記載の発明によれば、起点のノードからの階層数によって探索範囲を制限しているので、探索される

ノードは起点のノードと意味的なつながりの強いもののみとなる。また階層数により探索範囲を制限できるので、必要な範囲での情報のみを収集することができる。また、不要なノードの探索が行われないので、サーバおよびサーバとの間の通信回線の負担を軽減できるとともに探索に要する時間を短くすることができる。

【0093】また請求項4記載の発明によれば、1度読み込んだことのあるノードを再度読み込むことを防止したので、ループした範囲を繰り返し探索することを回避することができる。これにより探索を効率良く行うことができる。

【図面の簡単な説明】

【図1】本発明の一実施例におけるノード・リンク探索装置の構成の概要を表わしたブロック図である。

【図2】ノードテーブルの登録内容の一例を表わした説明図である。

【図3】リンクテーブルの登録内容の一例を表わした説明図である。

【図4】ノード・リンク探索装置の行う動作の流れを表わした流れ図である。

【図5】探索範囲内に存在するノードであるか否かを判定する際の処理の流れを表わした流れ図である。

【図6】探索範囲内に存在するノードであるか否かを判定する処理の他の一例の流れを表わした流れ図である。

【図7】図4に示したテーブル更新処理の流れを表わした流れ図である。

【図8】変形例のノード・リンク探索装置で用いられるノードテーブルの登録内容の一例を表わした説明図である。

【図9】変形例のノード・リンク探索装置で用いられるリンクテーブルの登録内容の一例を表わした説明図である。

る。

【図10】変形例におけるノード・リンク探索装置の行う動作の流れを表わした流れ図である。

【図11】ハイパーリンクが探索範囲内に存在するか否かを判定する際の処理の流れを表わした流れ図である。

【図12】ハイパーリンクが探索範囲内に存在するか否かを判定する処理の他の一例の流れを表わした流れ図である。

【図13】図10に示したテーブル更新処理の流れを表わした流れ図である。

【図14】ノードとこれらノード間を接続するハイパーリンクの一例を表わした説明図である。

【図15】ノードのドキュメント同士の関係の一例を表わした説明図である。

【図16】従来から使用されているノード・リンク探索装置の構成の概要を表わしたブロック図である。

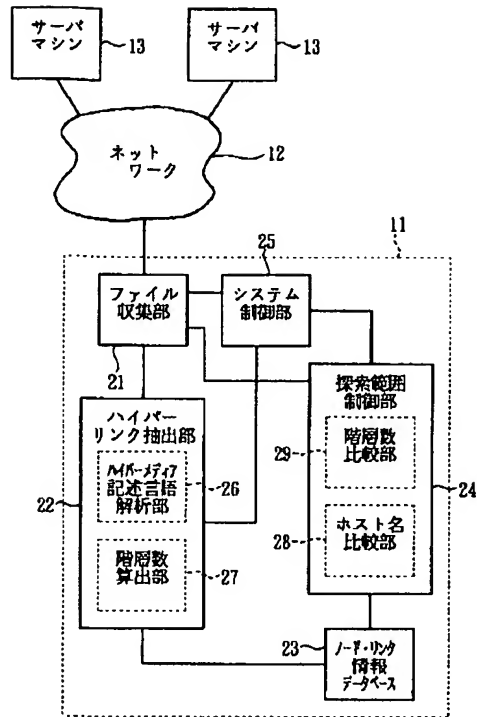
【符号の説明】

- 11 ノード・リンク探索装置
- 12 ネットワーク
- 13 サーバマシン
- 21 ファイル収集部
- 22 ハイパーリンク抽出部
- 23 ノード・リンク情報データベース
- 24 探索範囲制御部
- 25 システム制御部
- 26 ハイパーメディア記述言語解析部
- 27 階層数算出部
- 28 ホスト名比較部
- 29 階層数比較部
- 31、71 ノードテーブル
- 51、81 リンクテーブル

【図2】

	32 ノード 識別子	33 スキーム	34 ホスト 名	35 ポート 番号	36 スキーム特有部	37 階層 数	38 登録日時	39 探索日時	41 最終更新日時
43	1	http	host A	80	TOP.html	0	19951101224731	19951103193947	19951028135209
44	2	http	host B	80	Introduction.html	1	19951103193948		42
	m	http	host X	10080	Special.html	n			

【図1】

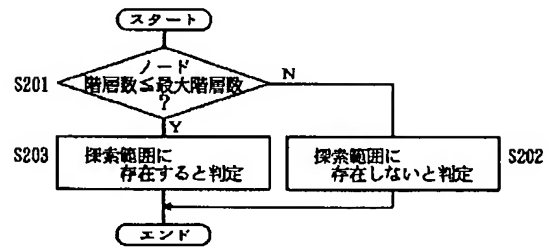


【図3】

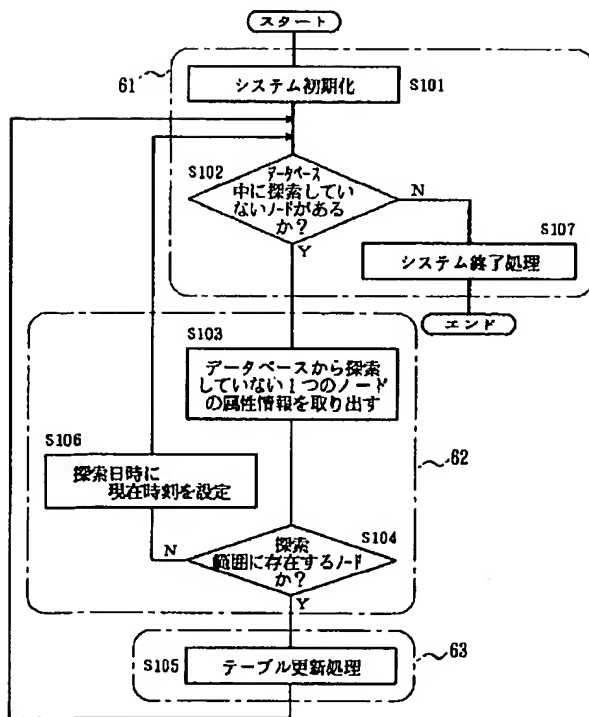
52 リンク識別子	53 リンク元 ノード識別子	54 リンク先 ノード識別子
1	1	2
2	1	3
56		
m	j	k

51

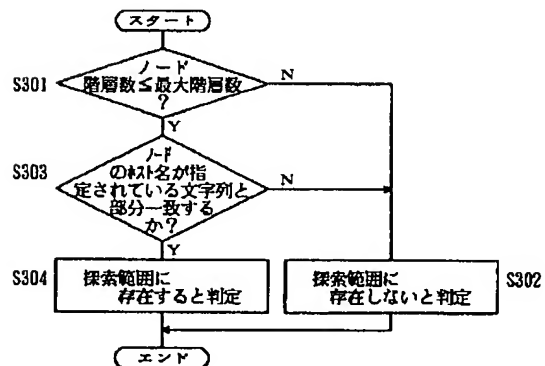
【図5】



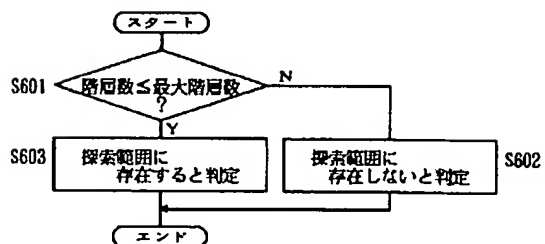
【図4】



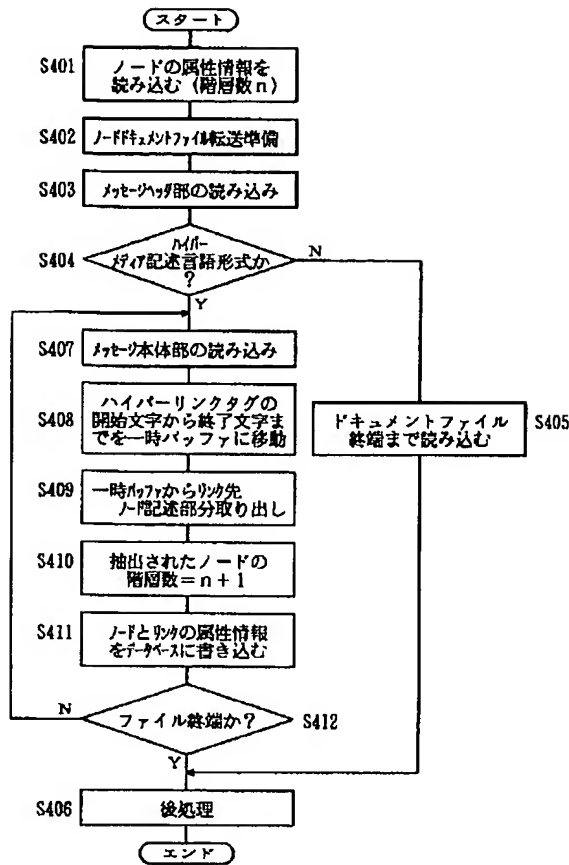
【図6】



【図11】



【図7】



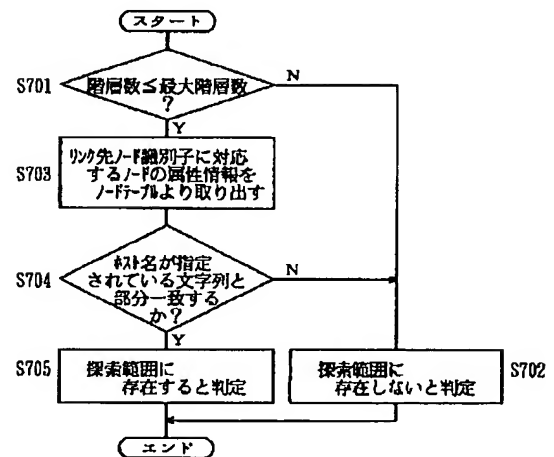
【図8】

32 ノード 識別子	33 スキーム	34 ホスト 名	35 ポート 番号	36 スキーム特有部	38 登録日時	41 最終更新日時
72 1	http	host A	80	TOC.html	19951101224731	19951028135209
74 2	http	host B	80	Introduction.html	19951103193948	73
m	http	host X	10080	Special.html		

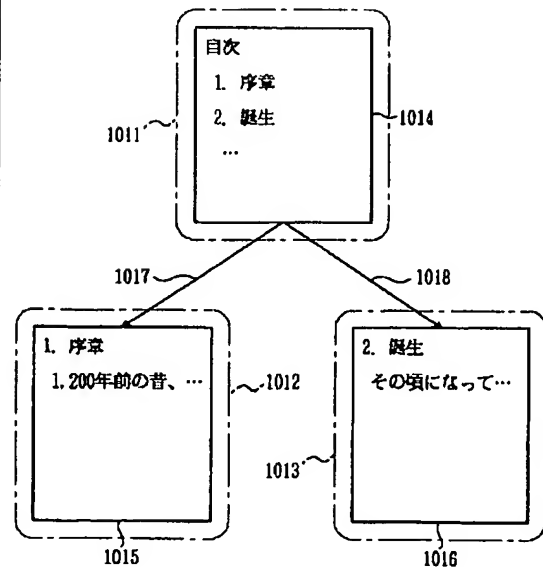
【図9】

52 リンク 識別子	53 リンク元 ノード識別子	54 リンク先 ノード識別子	82 階層 数	83 探索日時
84 1	0	1	0	19951103193947
86 2	1	2	1	85
92 3	1	3	1	
87 93	88	89		
m	j	k	33	

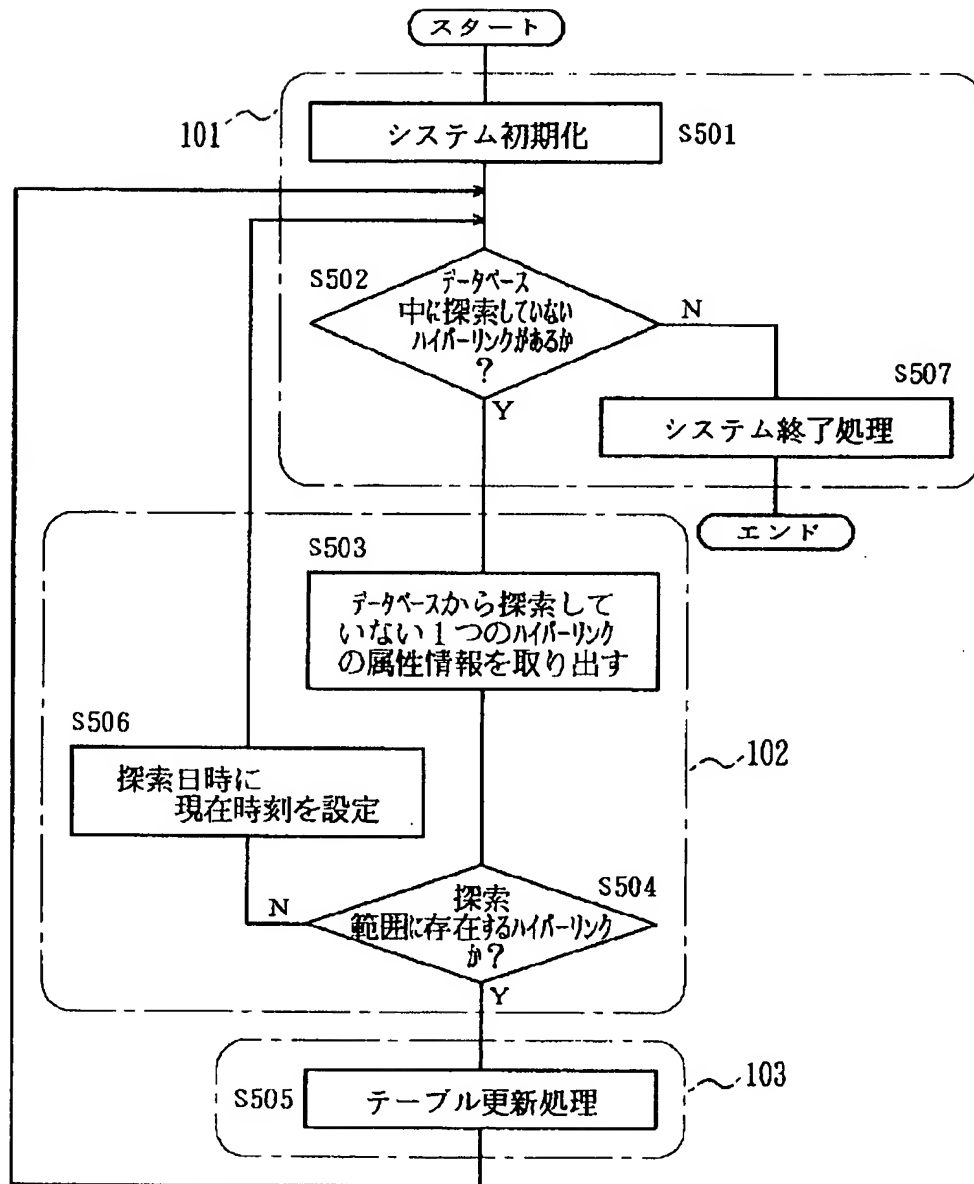
【図12】



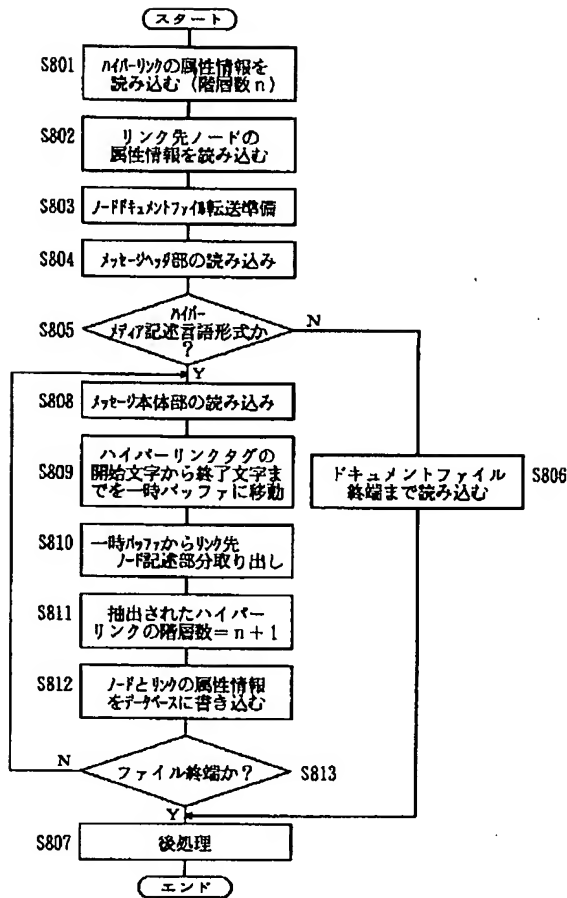
【図15】



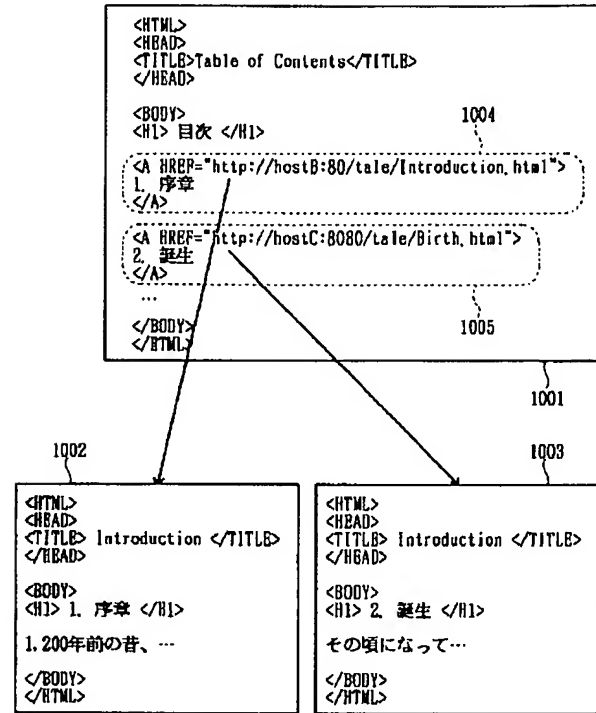
【図10】



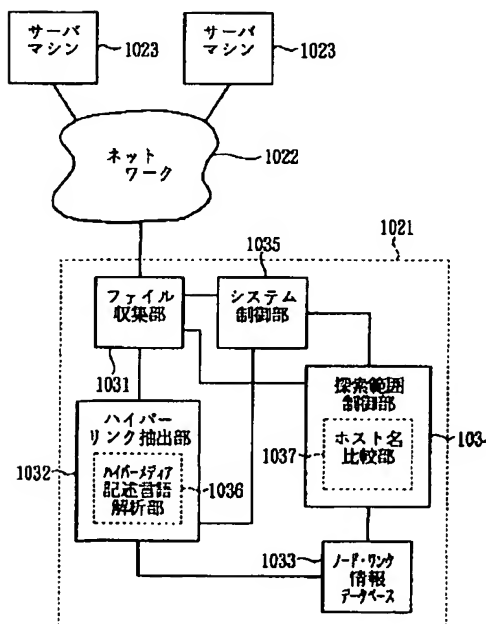
【図13】



【図14】



【図16】



* NOTICES *

JPO and INPIT are not responsible for any damages caused by the use of this translation.

1. This document has been translated by computer. So the translation may not reflect the original precisely.
2. **** shows the word which can not be translated.
3. In the drawings, any words are not translated.

DETAILED DESCRIPTION

[Detailed Description of the Invention]

[0001]

[Field of the Invention] This invention relates to the node links searching device which finds out the range of the node which should be made the object of the search in the bottom of the conditions specified in the database which consists of a node and a link. A node is related with the node links searching device used in the database distributed and stored in two or more servers on a network like especially a distributed hyper media system.

[0002]

[Description of the Prior Art] There is a hypertext system to which it refers by following the link with which these dependencies were expressed for two or more nodes which stored varieties of information, such as a document file, as one of the databases. Search by such a hypertext can be smoothly advanced by holding beforehand the index of the information stored in the node, and the link relation between nodes.

[0003] The hypertext system which can be displayed on JP, 4-321144, A so that the relation of the mutual between nodes can be grasped easily is indicated. Since each node can be linked to arbitrary nodes, a link may be formed in looped shape among two or more nodes. In this system, the display screen is created based on the table which registered the link relation between nodes beforehand. And the link which is looped shape is separated, and when it starts from a certain node, the link relation between nodes is made to be displayed as a tree structure.

[0004] Although the hypertext system was conventionally built to the local workstation etc., what distributes and stores a node in two or more servers connected via the network by development of communication technology in recent years has appeared. Such a system is called the distributed hyper media system. For example, World Wide Web (it expresses below World-Wide-Web WWW.) attracts attention as an information dispatching means on the Internet. Hypermedia is a hypertext which can treat not only text data, such as a character and a table, but the multimedia data of an animation, a sound, etc.

[0005] The table which expressed the link relation between nodes with the distributed hyper media system by acquiring the information on the node stored in each server via a network is created. The information showing the link relation between nodes will be called a node links information database. The device which creates a node links information database will be called a node links searching device.

[0006] In World Wide Web (WWW), the node distributes and exists in two or more

server machines on the Internet, Data of the node accumulated in each server It is transmitted according to the procedure called a hypertext transfer protocol (it expresses below HyperText Transfer ProtocolHTTP.). The multimedia document which is a node is described in the hypermedia description language form called a hypertext markup language (it expresses below HyperText Markup LanguageHTML.).

[0007]A node links searching device analyzes the contents of the node described in HTML acquired via the network, and extracts the hyperlink showing the storing position of the node of the next link destination. A hyperlink is searching the tag in which the predetermined character string (this is called the tag.) and end position which shows the starting position specified in HTML are shown, and is extracted as a character string between these.

[0008]The hyperlink showing a link destination node is described by the notation form called uniform resource ROKETAZU (it expresses below Uniform Resource LocatorsURL.). Search of a node and a hyperlink can be performed by finding out the portion put between the tags in which a hyperlink is shown out of the text of the acquired node, and extracting the character string described by URL from this inside. A node links searching device builds a node links information database by repeating such search. As a node links searching device, there are some which are called "WWW Wanderer", "WWW Robot", "WWW Spider", etc.

[0009]As for the hyperlink showing a link destinationnode, the following composition is defined in URL.

With <: [scheme (Scheme)>] <scheme characteristic part (Scheme-Specific-Part)> World Wide Web (WWW), as a transfer protocol of the document file of a node. In the case of the hypertext transfer pro col (HTTP) scheme used, a hyperlink is expressed as follows.

http://<host>:<port>:<path>?< searchpart> [0010]Here, "<host>" expresses the port number at the time of "<port>" communicating the host name of the server machine with which a document file exists. A port number is omissible, and when it omits, it is dealt with as what the standard value "80" was specified as. "<path>" expresses the place where the document file on a server machine exists. "< searchpart>" is a field used when passing data to a server machine.

[0011]Drawing 14 expresses an example of the hyperlink which connects between a node and these nodes. The document file contained in the node 1001, the node 1002, and the node 1003, respectively is described by HTML. The contents are classified according to various kinds of tags, and the contents of the document file are identifiable. Among these, the tag in which a hyperlink is shown is . -- It is expressed with and the portion corresponding to "URL" shows the link destination node. The fields 1004 and 1005 enclosed with a dotted line express the hyperlink with the document file of the node 1001, respectively.

[0012]The node 1001 presupposes that it is a node written like "http://hostA/tale/TOC.html" by URL. The nodes 1002 shall be "http://hostB:80 / tale/Introduction.html", and a node by which it writes "http://hostC:8080 / tale/Birth.html". [the node 1003] The hyperlink 1004 in the node 1001 is pointing to the node 1002. The hyperlink 1005 in the node 1001 is pointing to the node 1003. Thus, the node of a link destination is expressed by the hyperlink registered into the document file of a node.

[0013]Drawing 15 expresses an example of the relation of the documents of a node. The field enclosed with an one-point dashed line expresses the server in which the node is stored. the field 1011 -- a host name -- the server machine of "A" -- in the field 1012, a host name expresses the server machine of "B" and, as for the

field 1013, the host name expresses the server machine of "C", respectively.

[0014]The contents 1014-1016 which removed the tag etc. which were specified in HTML among the document files contained in the nodes 1001-1003 shown in drawing 14 are expressed in the field 1011-1013 showing storing, now the server which is of the node. The hyperlinks 1017 and 1018 express that the link destination nodes of the node 1001 are the node 1002 and the node 1003.

[0015]Next, the composition of the node links searching device which builds a node links information database based on the hyperlink registered into each node is explained.

[0016]Drawing 16 expresses the outline of the composition of the node links searching device currently used from the former. Two or more server machines 1023 are connected to the node links searching device 1021 via the networks 1022, such as the Internet. The node which stored the document file is accumulated in each server machine 1023. The node links searching device 1021 is provided with the file collecting part 1031 which collects the files of a node from the server machine 1023, and the hyperlink extraction part 1032 which extracts a hyperlink out of the file of the node acquired from the server machine 1023.

[0017]It has the node links information database 1033 which accumulates and manages various attribution information, such as a storage location of a node, and the link information between nodes based on the extracted hyperlink. The search range control section 1034 is a portion which restricts the search range of a node to the range corresponding to the conditions of the search range given from the input terminal which is not illustrated or the external file for setting out. The system control part 1035 is a circuit part which controls the flow of operation of each part in the node links searching device 1021 in generalization.

[0018]The hyperlink extraction part 1032 is provided with the hypermedia description language analyzing parts 1036 which analyze the contents of the multimedia document file of a node currently written with the hypermedia description language. The hypermedia description language analyzing parts 1036 extract a hyperlink and a link destination node by detecting various tags. A search range is specified by the host name as an identifier of the server machine 1023. The search range control section 1034 is provided with the host name comparing element 1037 which compares the host name as a search range specified as the host name of the server machine of the storage location of a node. By comparing the host name showing the storing position of a node as a part of information, it is judged whether a node is a thing in a search range.

[0019]The file collecting part 1031 of the node links searching device 1021 reads the multimedia document file of a node from the server machine 1023 which is in agreement with the host name specified as a search range through the network 1022. And the information on a hyperlink and a link destination node is extracted from the file read by the hyperlink extraction part 1032. The search range control section 1034 continues search to the link destination node, when the host name of the server machine with which the extracted link destination node is stored is in a search range. On the other hand, when the server machine of the storage location of a link destination node is outside a search range, the search to a previous node is stopped from it. The various attributes about the node and hyperlink in a search range are registered into the node links information database 1033.

[0020]Thus, the node links searching device which cannot specify at all a search range other than what specifies a host name as a search range also exists.

[0021]

[Problem(s) to be Solved by the Invention]The node links searching device which cannot specify a search range makes the whole network, such as the Internet, a search range, and searches for all the nodes which exist in all the server machines on a network. For this reason, originally the unnecessary information about nodes other than what should be made a search range was also accumulated in the node links information database, and the working efficiency at the time of it not only wasting those resources, but searching was also falling. Time to access the server machine with which a node exists, and a network becomes long, and there is also a problem of exploitation of other users' network resource being restricted and keeping.

[0022]Even if it is a device which can specify a search range, it can only perform restricting a search range by making a host name into a unit conventionally. This will have restricted the search range in the physical position of a node. However, since sequential retrieval of the hypertext is carried out based on the semantic relation between nodes, suitable range specification cannot be performed in specification of the search range which makes a server machine a unit.

[0023]For example, in the example shown in drawing 15, the node 1014 by which the table of contents was registered on the server "A" 1011 exists, and the nodes 1015 and 1016 by which the text of the contents corresponding to a table of contents was registered on the server "B" 1012 and the server "C" 1013 are registered, respectively. In such a case, servers "A", servers "B", and all the servers "C" must be specified as a host name of a search range to consider it as a search range to the contents of not only a table of contents but the document currently written. As a result, even many nodes of the table of contents which exists on a server "B" and a server "C", and the others which do not have semantic relation in any way serve as a search range, and there is a problem that search of an unnecessary node will be performed.

[0024]Then, the purpose of this invention is to provide the node links searching device which can restrict search of an unnecessary node without semantic relation.

[0025]

[Means for Solving the Problem]In the invention according to claim 1. It carries out based on link information showing a name of a node of a link destination and a storing position of a node of a link destination which are included in each node of a hypertext. A search condition setting-out means to set up as the maximum of a hierarchy number which is the number of nodes which exist in from a node used as a starting point of search of a limited condition of a search range at the time of performing search to arbitrary nodes of a link destination from a node one by one before a node of a search place, A file collecting means which performs repeatedly reading the contents of the node of a link destination which link information shows from a server which is accumulating it sequentially from a node used as a starting point of search, A node links information storage means which memorizes information showing matching with a node of a link destination which link information which that node includes, and this show, and a node read this time whenever it reads the contents of one node by this file collecting means, A hierarchy number calculating means which asks for a hierarchy number of a node of a link destination which link information included in the node shows whenever it reads the contents of one node by a file collecting means, A hierarchy number for which it asked by this hierarchy number calculating means by a search condition setting-out means. When larger than the maximum of a set-up hierarchy number, a node links searching device is made to possess a search range limit means which stops reading by a file collecting means

of the contents of the node linked after a node read this time.

[0026]That is, in the invention according to claim 1, the number of nodes which exist in from a node used as a starting point of search before a node of a search place is made into a hierarchy number, and conditions which restrict a range which searches are specified by the maximum of a hierarchy number. By following a link from a node used as a starting point, the contents of the node are read from a server one by one, and correspondence relation between the name and storing position and a linking agency node, and a link destination node is registered about a node of a link destination of a read node. It asks for a hierarchy number about a node of a link destination of a node read whenever it read a node from a server. When it becomes more than the maximum which this hierarchy number set up, it stops following a link to the point more and reading a node.

[0027]The range of search is restricted with a hierarchy number from a node made into a starting point by these. Thus, since a hierarchy number from a node of a starting point has restricted a search range, a node for which it is searched becomes only a strong thing of as semantic relation as a node of a starting point. Since a hierarchy number can restrict a search range, only information on a required range is collectable.

[0028]In the invention according to claim 2. It carries out based on link information showing a name of a node of a link destination and a storing position of a node of a link destination which are included in each node of a hypertext. A search condition setting-out means to set up as the maximum of a hierarchy number which is the number of links which exist in from a node used as a starting point of search of a limited condition of a search range at the time of performing search to arbitrary nodes of a link destination from a node one by one before a node of a search place, A file collecting means which performs repeatedly reading the contents of the node of a link destination which link information shows from a server which is accumulating it sequentially from a node used as a starting point of search, A node links information storage means which memorizes information showing matching with a node of a link destination which link information which that node includes, and this show, and a node read this time whenever it reads the contents of one node by this file collecting means, A hierarchy number calculating means which asks for a hierarchy number of a link in a node of a link destination which link information included in the node expresses whenever it reads the contents of one node by a file collecting means, A hierarchy number for which it asked by this hierarchy number calculating means by a search condition setting-out means. When larger than the maximum of a set-up hierarchy number, a node links searching device is made to possess a search range limit means which stops reading by a file collecting means of the contents of the node linked after a node read this time.

[0029]That is, in the invention according to claim 2, the number of links which exist in from a node used as a starting point of search before a node of a search place was made into a hierarchy number, and the maximum of a hierarchy number has restricted a range which searches. Thereby, only information about a node of a starting point and a node with strong semantic relation is collectable. The name and storing position, and link relation between nodes are collectable only about a node in a search range specified with a hierarchy number.

[0030]A node which is the target of search is distributed and stored in two or more servers connected to a network in the invention according to claim 3.

[0031]That is, each node which is the target of search is distributed and stored in

two or more servers connected via a network in the invention according to claim 3. Since a hierarchy number has restricted a search range, even if a node is distributed by two or more servers, only information on a required node is collectable from many servers. For example, when a search range can be specified only per server, it may be collected to a node of not only a node that exceeds a required hierarchy number from a node made into a starting point but a starting point, and an unrelated node without link relation. Even if a node of a search range is distributed by two or more servers by restricting a search range with a hierarchy number, only information about a node of a required range which has semantic relation from a node made into a starting point is collectable.

[0032]In the invention according to claim 4, a file collecting means, The same node discriminating means that distinguishes whether a node of a link destination which link information included in a node read this time shows is the same as that of an already read node, When it is distinguished that it is the same as that of a node already read by this same node discriminating means, a multiplex reading stop means to stop reading of that node for the second time is provided.

[0033]That is, in the invention according to claim 4, reading of a node read once for the second time is prevented. It is avoidable that this repeats and searches for a range which carried out the loop.

[0034]

[Embodiment of the Invention]

[0035]

[Example]Drawing 1 expresses the outline of the composition of the node links searching device in one example of this invention. The search procedure on hypermedia structure can be regarded as the tree structure which uses as a root the node which started search. Then, the depth of the tree which has each node by making a root into the starting point is made into a hierarchy number, and a search range is restricted with the hierarchy number from the starting point.

[0036]Two or more server machines 13 which accumulated the node via the networks 12, such as the Internet, are connected to the node links searching device 11. The node links searching device 11 is provided with the hyperlink extraction part 22 which extracts a hyperlink and a link destination node out of the file of the file collecting part 21 which collects the files of a node from the server machine 13, and the node acquired from the server machine 13.

[0037]The node links searching device 11 has the node links information database 23 which accumulates and manages the various attribution information of the extracted hyperlink and a link destination node. The search range control section 24 is a portion which restricts the search range of a node so that it may agree on the conditions for specifying the search range inputted from the file the input terminal which is not illustrated or for setting out. The system control part 25 is a circuit part which controls the flow of operation of each part in the node links searching device 11 in generalization.

[0038]The hyperlink extraction part 22 is provided with the hypermedia description language analyzing parts 26 which analyze the contents of the multimedia document file of a node written with the hypermedia description language. The hypermedia description language analyzing parts 26 extract a hyperlink and a link destination node by detecting various tags. The hyperlink extraction part 22 is provided with the hierarchy number calculation part 27 which computes the hierarchy number of the extracted link destination node. The hierarchy number calculation part 27 asks for what added "1" to the hierarchy number of the node of a linking agency as a

hierarchy number of a link destination node.

[0039]A search range is specified as the host name as an identifier of the server machine 13 with the maximum hierarchy number. The search range control section 24 is provided with the hierarchy number comparing element 29 which compares with the hierarchy number of the node and the maximum hierarchy number as a limit value of a search range the host name comparing element 28 which compares the host name of the storage location of a node with the host name specified as a search range. The search range control section 24 restricts a search range by the hierarchy number of a node, and the host name of a storage location. In order to prevent multiplex reading of the same node, it also performs restricting the range of search by the lapsed time from the last search.

[0040]The file collecting part 21 reads the multimedia document file of a node from the server machine 13 through the network 12. The hyperlink extraction part 22 extracts a hyperlink and a link destination node from a multimedia document file based on the analysis result of the hypermedia description language analyzing parts 26. Next, the hierarchy number calculation part 27 computes the hierarchy number of a link destination node or a hyperlink by adding "1" to a linking agency node.

[0041]The search range control section 24 restricts a search range based on whether it being below the maximum hierarchy number as which the hierarchy number of the node or the hyperlink was specified, or a host name should be match-partial-specified. About the node and hyperlink in a search range, those attribution information is accumulated in the node links information database 23. The table which registers the attribute about a node within the node links information database 23 will be called a node table, and the table which registers the attribute about a link will be called a link table. Thus, collection of the attribution information about an unnecessary node and the registration to a database are prevented with restricting a search range with a hierarchy number.

[0042]Drawing 2 expresses an example of the contents of registration of a node table. the node table 31 -- the node identifier 32 from the left of a figure, scheme (Scheme)33, the host name 34, the port number 35, and a scheme characteristic part (Scheme-Specific) -- 41 is registered at 38, the search time 39, and the time of a last update date at the time of 36, the hierarchy number 37, and a registration date. Among these, in drawing 1, when the node identifier 32 registers the attribution information of a node into a node table with the data base management system (DBMS) which is not shown, he is an identifier generated automatically. The scheme 33, the host name 34, the port number 35, and the scheme characteristic part 36 correspond to each portion of the character string of URL which shows the storing position of a node. The hierarchy number 37 is an item showing the hierarchy number of the node from a search starting position.

[0043]38 shows the time at which the node was first registered into the database at the time of a registration date. The search time 39 expresses the newest time searched for this node. 41 expresses the time of the last update date of the multimedia document file of a node at the time of a last update date. The item of 41 is an attribute updated when registered as a node of a linking agency among a figure at the search time 39 enclosed with the dotted line 42, and the time of a last update date.

The item enclosed with the dotted line 43 is an attribute updated when registered as a link destination node to which the hyperlink of the linking agency node points.

[0044]Drawing 3 expresses an example of the contents of registration of a link table. As for the link table 51, the link identifier 52, the linking agency node identifier 53, and the link destination node identifier 54 are registered from the left of a figure. The identification number of the node automatically assigned by DBMS is registered into these items. The node identifier of the node of the point which a hyperlink points [the node identifier of the node of the source origin of the hyperlink shown by the link identifier 52] out to the linking agency node identifier 53 is registered as the link destination node identifier 54.

[0045]The information on the node used as the starting point of search is beforehand registered into the node table 31 of the node links information database at least one. Under the present circumstances, the item which should be registered is 38 at the time of the node identifier 32, the scheme 33, the host name 34, the port number 35, the scheme characteristic part 36, the hierarchy number 37, and a registration date. "0" is set to the hierarchy number of the node used as the starting point of search.

[0046]Drawing 4 expresses the flow of the operation which a node links searching device performs. Processing of the step enclosed with the one-point dashed line 61 is performed by the system control part 25 of a node links device. Processing of the step which surrounded processing of the step enclosed with the one-point dashed line 62 with the one-point dashed line 63 by the search range control section 24 is performed by the file collecting part 21 and the hyperlink extraction part 22.

First, in advance of search, a system control part initializes systems, such as an input of a search condition (Step S101). A search condition is inputted from the input terminal and external file which are not shown by drawing 1. As a search condition, the maximum (the maximum hierarchy number) of the hierarchy number of the node which should be made a search range here, the lapsed time for judging the last search results to be an old thing, and the character string (a part of host name to search or all) for specifying the server machine made into a search object are set up as conditions.

[0047]The system control part 25 investigates whether the node table 31 is searched and an unsearched node exists (Step S102). When the search time 39 registered into the node table 31 is an undefined (unregistered), and it corresponds to either in the case of beyond the lapsed time to which the search time 39 was set as a search condition rather than current time being the past time, it judges with an unsearched node. When an unsearched node exists, (Step S102; Y) and the search range control section 24 take out the attribution information of one node for which it is not searched from the node table 31. Under the present circumstances, a node is preferentially chosen from what has a small hierarchy number (Step S103). Therefore, the attribution information of the node of the search starting point whose hierarchy number is "0" is taken out at first.

[0048]Next, it is distinguished whether it is a node which exists in the search range where the selected node was specified (Step S104). The processing which distinguishes whether it is inside of a search range is explained in detail later. In being a node which exists in a search range, it reads the document file of (Step S104; Y) and its node through a network, and a table information update process (Step S105), such as carrying out additional registration of the link destination node described by this to a node table, is performed. The details of a table update process are explained later. After finishing the update process (Step S105) about one node, it returns to Step S102 again, and the search processing about an unsearched node is repeated.

[0049]In being a node outside the search range where the selected node was specified, it changes the search time 39 about the selected node in (Step S104; N) and the node table 31 at current time (Step S106), and it returns to Step S102. When an unsearched node stops existing in a node table (step S102;N), end processing for ending search is performed (Step S107), and processing is ended (end).

[0050]Drawing 5 expresses the flow of the processing at the time of judging whether it is a node which exists in a search range. Here, it is distinguished only with the hierarchy number whether it is inside of a search range. The search range control section 24 compares first the maximum hierarchy number set up as the hierarchy number and search condition of the selected node (Step S201). When the hierarchy number of a node is larger than the maximum hierarchy number, (Step S201; N) and this node judge with not existing in a search range (Step S202). The hierarchy number of a node judges with the thing which has the good maximum hierarchy number and to which it is small, or (Step S201; Y) and this node exist in a search range in being equal (Step S203).

[0051]Drawing 6 expresses the flow of other examples of processing which judge whether it is a node which exists in a search range. Here, distinction by a host name is added besides distinction by a hierarchy number. The search range control section 24 compares the selected hierarchy number and the maximum hierarchy number of a node (Step S301), and when the hierarchy number of a node is larger than the maximum hierarchy number, (Step S301; N) and this node judge with not existing in a search range (Step S302). It is investigated whether the character string with the maximum hierarchy number sufficient [the hierarchy number of a node] which was small, or was specified as a host name of (Step S301; Y) and a search range when equal, and the host name of a node carry out match partial (Step S303).

[0052]For example, when "AB" is specified as a search string, all the things containing the character string "AB" specified from the head of host names, such as "ABC" and "ABD", will be distinguished if match partial is carried out. When match partial has not been carried out like "CAB", it judges that (Step S303; N) and this node do not exist in a search range (Step S302). In carrying out match partial, (Step S303; Y) and this node judge with existing in a search range (Step S304). Here, after judging the hierarchy number, the match partial of the host name was judged, but it may carry out by replacing these order.

[0053]Drawing 7 expresses the flow of the table update process shown in drawing 4. This processing is performed by the file collecting part 21 and the hyperlink extraction part 22. First, it performs reading the contents of the node distinguished when it existed in the search range by Step S104 of drawing 4 from the server machine with which it is stored through the network. For this reason, the file collecting part 21 reads the attribution information of an applicable node from the node table 31 of the node links information database 23 (Step S401). The hierarchy number of the read node is temporarily set to n here. Next, the file collecting part 21 makes connection with the server machine 13, after initializing the communication buffer and temporary buffer which are used for communication with a server machine and which are not illustrated, and it makes transmission (reading) preparations of the document file of a node (Step S402).

[0054]Then, the file collecting part 21 starts transmission of a document file according to the transfer protocol of a distributed hyper media system. The file collecting part 21 reads a transfer protocol message header part (Step S403), and investigates the format of the description language of a document file, etc. (Step

S404). When the read file is not hypermedia description language form, it reads to (Step S404; N) and the termination of the document file (Step S405). And post-processing of making release of a communication buffer and a temporary buffer and connection with a server machine is performed (Step S406), and processing is ended (end).

[0055]When the read file is hypermedia description language form, (Step S404; Y) and the file collecting part 21 read the DOKYUYUMENTO file of a transmission protocol message main part into a communication buffer (Step S407). Next, from the initial-statement character of the tag showing the portion of a hyperlink to a trailing character is moved to a temporary buffer by the hyper-MEDIADO description language analyzing parts 26 (Step S408). The portion which has described the link destination node is taken out out of the hyperlink stored in the temporary buffer (Step S409). A hierarchy number calculation part calculates "n+1" which added "1" to the hierarchy number "n" of the node of a linking agency as a hierarchy number of a link destination node (Step S410).

[0056]Thus, the attribution information of the obtained node and a hyperlink is written in the node table 31 and the link table 51 in the node links information database 23 (Step S411). Under the present circumstances, the one-copy each item of the attribution information of a linking agency node and the attribution information of a link destination node and all the items of the attribution information (link table) of a hyperlink are written in. It means carrying out one additional registration of the attribution information and the hyperlink from a node to the read link destination node about the link destination node which one hierarchy linked to the node read this time by this followed to each table. However, when already registered about the same node, additional registration to a table is not performed but performs only renewal of search time etc. Thereby, the multiple registration of the same node or a hyperlink is avoided.

[0057]Next, it investigates whether it processed to the termination of the DOKYUYUMENTO file of the node read this time (Step S412), and when not resulting in a termination, it returns to (Step S412; N) and Step S407. Since each node may have two or more link destination nodes, it performs registration to a node table and a link table about all the link destination nodes of the node read this time by repeating such processing and performing it to the termination of a file. When it processes to the termination of a file, (Step S412; Y) and the table update process about this node are ended (end).

[0058]Here, the contents of the attribution information registered into a node table and a link table in Step S411 of drawing 7 are explained. Make into a linking agency node the node read from the server machine, and let the node described by the hyperlink contained in this node be a link destination node. First, about a linking agency node, 41 is registered at the search time 39 and the time of a last update date. Thereby, it can leave the newest time information of when to have searched the node. For example, in the node table 31 shown in drawing 2, the node identifier should read the node of "1" from the server machine, and the node of "2" should be described for the node identifier as the point which the hyperlink of this points out. In this case, a node identifier is [the node of "1"] a linking agency node, and a node identifier is [the node of "2"] a link destination node.

[0059]The search time surrounded by the dotted line 42 in drawing 2 and the time of a last update date are updated as attribution information of a linking agency node. The portion surrounded by the dotted line 44 of drawing 2 is registered as attribution information of a link destination node. Namely, it is at the

registration date time which are the hierarchy number corresponding to the character string of URL which indicates the storing position of a node to be a node identifier for which it asked by the scheme, the host name, the port number, the scheme characteristic part, and the hierarchy number calculation part, and the time registered into the database. That by which the node identifier was generated automatically by DBMS is registered. Since the hyperlink which actually reads from the server machine 13 still, and is contained in it about the node of a link destination is not investigated, it becomes unregistered with as at search time and the time of a last update date.

[0060]The link identifier 52 generated automatically by DBMS, the linking agency node identifier 53, and the link destination node identifier 54 are registered into the link table 51. Drawing 3 is explained to an example. A node identifier reads the node of "1" and a node identifier assumes that the node of "2" is a link destination node by the hyperlink of this. First, since the node identifier of the node of the source origin of the hyperlink (55) of "1" is "1", a link identifier registers this value as a linking agency node identifier (56). Since the node identifier of the node of the point which a hyperlink points out is "2", he registers this value as a link destination node identifier (57). When other hyperlinks of the read node are registered, a link identifier is assigned also about these links and a linking agency node identifier and a link destination node identifier are registered. When 2 hyperlinks exist, the information on the range shown by the dotted line 58 of drawing 3 is registered into the link table 51.

[0061]The renewal of the node table 31 and the link table 51 reads one node from the server machine 13, and whenever it finds out a new link destination, it is performed. Therefore, if it returns to Step S102 after finishing a table update process (Step S105) about one node in the flow chart of drawing 4, the node newly registered in the table update process will also become the search object. Thereby, search to a link destination is advanced one after another. However, the search range is restricted by conditions, such as a hierarchy number.

[0062]Since the time near current time is registered into the search time of the node which makes it the unsearched judging standard for search time to be the past time in the undefined or beyond lapsed time and for which it searched once, investigation of a link destination is not conducted in piles about the same node. Thereby, by a node "B", even if the link destination of a node "A" is a case where the loop is formed by a link like a node "A" in the link destination of a node "B", it does not investigate a node "A" again, and does not repeat and follow a loop to the maximum hierarchy number, for example.

[0063]Thus, since the hierarchy number of a node has restricted the range of search, it can search only for the node used as a starting point, and the strong node of semantic relation. Unnecessary search can be reduced compared with the case where a search range can be restricted only per server.

[0064]Modification [0065]Although the hierarchy number is given as attribution information of a node in the example described until now, it is distinguished in the modification whether it is inside of a search range based on the hierarchy number of a hyperlink. The composition of a device is the same as that of what was shown in drawing 1, and omits the explanation.

[0066]Drawing 8 expresses an example of the contents of registration of the node table used with the node links searching device of a modification. The same numerals are given to the same item as drawing 2, and those explanation is omitted suitably. The node table 71 is different from a hierarchy number in that the item

which registers search time is deleted compared with the node table 31 shown in drawing 2.

[0067]Drawing 9 expresses an example of the contents of registration of the link table used with the node links searching device of a modification. The same numerals are given to the same item as drawing 3, and those explanation is omitted suitably. In addition to the link table 51 shown in drawing 3, the link table 81 is provided with the item 82 of a hierarchy number, and the item 83 of search time. These are registered into the node table in the example. In advance of search, it is necessary to register the hyperlink used as the starting point of search into the at least one link table 81 in a modification. The hyperlink used as the search starting point registered here is virtual.

A linking agency node does not exist (unknown).

The attribution information about the link destination node which the hyperlink used as the starting point of search shows must be beforehand registered into the node table 71.

[0068]There is a node identifier of the link destination node which the hyperlink of the starting point of search shows first as contents registered into the node table 71 in advance of search. This value is automatically assigned by DBMS. The scheme, the host name, the port number, and the scheme characteristic part are registered the first stage as an item corresponding to the character string of URL which shows the storing position of a node. The time of the registration date when this node expresses the time registered into the database is registered the first stage. In the example of drawing 8, the item of the range enclosed with the dotted line 72 is registered the first stage.

[0069]There is a link identifier which is an identifier of the hyperlink used as the starting point of search as contents registered into the link table 81 the first stage. This value is generated automatically by DBMS. Since the link origin of the hyperlink used as the starting point of search is unknown, a linking agency node identifier's initial value is set to "0" showing the undefined. The link destination node identifier registers the same value as the node identifier whom DBMS assigned to the node used as a link destination. A hierarchy number registers being [it / the starting point of search] "0" the first stage. Search time registers the time which performed search about the hyperlink concerned.

As an initial value of a search start, it is considered as as [undefined].

In the example of drawing 9, the item enclosed with the dotted line 84 is registered the first stage.

[0070]Drawing 10 expresses the flow of the processing which the node links searching device in a modification performs. The step surrounded with the one-point dashed line 101 expresses the processing which the system control part 25 performs. The step by which the step surrounded with the one-point dashed line 102 was surrounded by the search range control section 24 with the one-point dashed line 103 expresses the processing performed by the file collecting part 21 and the hyperlink extraction part 22.

[0071]First, the system control part 25 initializes the system concerned (Step S501). Under the present circumstances, the maximum hierarchy number as the maximum of the hierarchy number of the hyperlink made into a search range and the lapsed time used as the standard which treats the last search results as an old thing are set up. In setting up the range of a server by a host name furthermore, it inputs the character string for restricting a host name. These are incorporated from the input terminal or external file which is not illustrated. Next, it is investigated

whether the hyperlink for which it is not searching in the link table 81 exists (Step S502). Search time is a thing of the undefined (unregistered) and the hyperlink for which it is not searched has [beyond lapsed time] search time older than current time.

[0072]When the hyperlink for which it is not searched exists in the link table 81, one of (Step S502; Y) and the hyperlinks for which it is not searched is chosen, and the attribution information is taken out (Step S503). Next, the search range control section 24 judges whether one selected hyperlink is in a search range (Step S504). Although judged by the both sides of a hierarchy number or a hierarchy number, and a host name, the details are explained later.

[0073]In existing in a search range, it performs (Step S504; Y) and a table update process (Step S505). In this processing, the attribution information of the link destination node which a hyperlink shows is taken out from the node table 71, the document file of this node is read from a server machine, and it performs carrying out additional registration of the attribution information of the link destination node registered into this etc. It mentions detailed flowing into processing later. After finishing the table update process (Step S505) about the link destination node which a hyperlink shows, it returns to Step S502 again, and the search processing about an unsearched hyperlink is repeated.

[0074]When the selected hyperlink is outside a search range, the search time 39 about the selected hyperlink in (Step S504; N) and the link table 81 is changed at current time (Step S506), and it returns to Step S502. When an unsearched hyperlink stops existing in the link table 81 (step S502;N), end processing for ending search is performed (Step S507), and processing is ended (end).

[0075]Drawing 11 expresses the flow of the processing at the time of judging ***** [a hyperlink / in a search range]. Here, it is distinguished only with the hierarchy number of a hyperlink whether it is inside of a search range. First, the search range control section 24 compares the maximum hierarchy number set up as the hierarchy number and search condition of the selected hyperlink (Step S601). When the hierarchy number of a hyperlink is larger than the maximum hierarchy number, (Step S601; N) and this hyperlink judge with not existing in a search range (Step S602). The hierarchy number of a hyperlink judges with the thing which has the good maximum hierarchy number and to which it is small, or (Step S601; Y) and this hyperlink exist in a search range in being equal (Step S603).

[0076]Drawing 12 expresses the flow of other examples of processing which judge ***** [a hyperlink / in a search range]. Here, distinction by a host name is added besides distinction by the hierarchy number of a hyperlink. The search range control section 24 compares the selected hierarchy number and the maximum hierarchy number of a hyperlink (Step S701), When the hierarchy number of a hyperlink is larger than the maximum hierarchy number, (Step S701; N) and this hyperlink judge with not existing in a search range (Step S702). The hierarchy number of a hyperlink takes out the attribution information about the link destination node with the sufficient maximum hierarchy number which it is small, or (Step S701; Y) and this hyperlink show in being equal from the node table 71 (Step S703).

[0077]The host name contained in the attribution information of the taken-out link destination node investigates whether match partial is carried out to the character string specified as a host name of a search range (Step S704). In carrying out match partial, the hyperlink of (Step S704; Y) and the point judges with existing in a search range (Step S705). In not carrying out match partial, it judges with (Step S705; N) and a hyperlink not existing in a search range (Step S702). Here,

after judging the hierarchy number, the match partial of the host name was judged, but it may carry out by replacing these order.

[0078]Drawing 13 expresses the flow of the table update process shown in drawing 10. This processing is performed by the file collecting part 21 and the hyperlink extraction part 22. First, the attribution information of the hyperlink distinguished when it existed in the search range by Step S504 of drawing 10 is read from the link table 81 (Step S801). The hierarchy number of the read hyperlink is temporarily set to n here. Next, the attribution information of the node corresponding to the link destination node identifier of this hyperlink is read from the node table 71 (Step S802).

[0079]The file collecting part 21 makes connection with a server machine, after initializing BAFFAKU the communication buffer which is used for communication with a server machine and which is not illustrated, and temporarily, and it makes reading preparations of the document file of the link destination node which a hyperlink shows (Step S803). Then, the file collecting part 21 starts transmission of a document file according to the transfer protocol of a distributed hyper media system.

[0080]The file collecting part 21 reads a transfer protocol message header part (Step S804), When the description language of a document file is not hypermedia description language form, it reads to (Step S805; N) and the termination of the document file (Step S806). And post-processing of making release of a communication buffer and a temporary buffer and connection with a server machine is performed (Step S807), and processing is ended (end).

[0081]When the read file is hypermedia description language form, (Step S805; Y) and the file collecting part 21 read the DOKYUYUMENTO file of a transmission protocol message main part into a communication buffer (Step S808). Next, from the initial-statement character of the tag showing the portion of a hyperlink to a trailing character is moved to a temporary buffer (Step S809), and the portion which has described the link destination node is taken out out of the hyperlink stored in the temporary buffer (Step S810). The hierarchy number calculation part 27 sets up "n+1" as a hierarchy number of the extracted hyperlink (Step S812).

[0082]Thus, the attribution information of the obtained node and a hyperlink is written in the node table 71 and the link table 81 in the node links information database 23 (Step S812). Under the present circumstances, the one-copy each item of the attribution information of a linking agency node and the attribution information of a link destination node and all the items of the attribution information of a hyperlink are written in. Thereby, registration is performed about the attribution information and the hyperlink from a node to the read link destination node about the link destination node of this read node. However, when the hyperlink which has a link destination the same link origin is already registered, additional registration to a table is not performed but performs only renewal of search time etc. It is avoided that multiple registration of the same node and hyperlink is carried out by this.

[0083]Next, it investigates whether it processed to the termination of the DOKYUYUMENTO file of the node read this time (Step S813), and when not resulting in a termination, it returns to (Step S813; N) and Step S808. Since each node may have two or more link destination nodes, registration to a node table and a link table is performed about all the link destination nodes of the node read this time by repeating such processing and performing it to the termination of a file. Processing is ended when it processes to the termination of a file (step S813;Y)

(end).

[0084]Here, the contents of the attribute registered into the node table 71 and the link table 81 in Step S812 of drawing 13 are explained. Make into a linking agency node the node read from the server machine, and let the node described by the hyperlink contained in this node be a link destination node. First, about a linking agency node, 41 is registered at the time of a last update date. For example, in the node table 71 shown in drawing 8, the node identifier should read the node of "1" from the server machine, and the node of "2" should be described for the node identifier as the point which the hyperlink of this points out. In this case, a node identifier is [the node of "1"] a linking agency node, and, in a node identifier, the node of "2" turns into a link destination node.

[0085]The time of the last update date surrounded by the dotted line 73 in drawing 8 is updated as attribution information of a linking agency node. The portion surrounded by the dotted line 74 of drawing 8 is registered as attribution information of a link destination node. Namely, it is at the registration date time which is the scheme, the host name, the port number, the scheme characteristic part, and the time registered into the database corresponding to the character string of URL which indicates the position of a node to be a node identifier. That by which the node identifier was generated automatically by DBMS is registered.

[0086]The link identifier generated automatically by DBMS, a linking agency node identifier, a link destination node identifier, a hierarchy number, and search time are registered into the link table 81. Drawing 9 is explained to an example. The node identifier whose link identifier is a link destination based on the hyperlink of "1" should read the node of "1." The node of "3" shall be linked [the node identifier] for "2" and a node identifier by the hyperlink described by the read node.

[0087]Under the present circumstances, since it searched with reference to the hyperlink 84, that time is registered as the search time 85. Next, the link identifier about the newly obtained hyperlink is acquired from DBMS, and it registers with the column of the link identifier which carries out additional registration of this. For example, a link identifier registers "2" as a link identifier (87) of the hyperlink (86) of "2." This value is an ID number suitably given by DBMS on the basis of the order of registration of a table, etc.

[0088]Since the node identifier's value is "1", the linking agency node of the hyperlink 86 registers "1" as the linking agency node identifier 88. When registering the attribute of the node of a link destination into the node table 71, a node identifier's value given by DBMS, i.e., "2", is registered into the link destination node identifier 89. Since it is the hyperlink acquired from the node which followed and obtained the hyperlink of the hierarchy number "0", the value "1" which added "1" to "0" is registered into the hierarchy number 91. It registers with the hyperlink 92 which a node identifier turns to the node of "3" by the same procedure as ** just. The information on the range shown by the dotted line 93 of drawing 9 about the hyperlink which carried out additional registration will be registered.

[0089]The renewal of the node table 71 and the link table 81 reads the link destination node from Sir BAMASHI based on one hyperlink, and whenever it finds out a new link destination, it is performed. Therefore, if it returns to Step S502 after finishing a table update process (Step S505) about one node in the flow chart of drawing 10, the hyperlink newly registered in the table update process will also become the search object. Thereby, search to a link destination is advanced one

after another. However, the search range is restricted by conditions, such as a hierarchy number.

[0090] Since it is making for search time to be the past time in the undefined or beyond lapsed time into the unsearched judging standard, it does not investigate again about the hyperlink for which it searched once. By this, even if the link destination of a node "A" is a case where the loop is formed by a link like a node "A" in the link destination of a node "B" by the node "B", for example, The hyperlink from a node "A" to a node "B" is not investigated again, and a loop is not followed to the maximum hierarchy number.

[0091] Thus, even if it restricts a search range based on the hierarchy number of a hyperlink, the same effect as the time of restricting a search range based on the hierarchy number of a node can be acquired. However, in giving a hierarchy number to a hyperlink, in order to investigate the link destination node of the hyperlink obtained with reference to the link table, it is necessary to also refer to a node table, and the reference process of a table increases to an example compared with the case where a hierarchy number is given to a node like.

[0092]

[Effect of the Invention] Since the hierarchy number from the node of a starting point has restricted the search range according to claim 1 thru/or the invention according to claim 3 as explained to details above, the node for which it is searched becomes only a strong thing of as semantic relation as the node of a starting point. Since a hierarchy number can restrict a search range, only the information on the required range is collectable. Since search of an unnecessary node is not performed, while the burden of the communication line between servers is mitigable, time which search takes can be shortened.

[0093] According to the invention according to claim 4, since it prevented reading again the node read once, it is avoidable to repeat and search for the range which carried out the loop. Thereby, it can search efficiently.

[Translation done.]

* NOTICES *

JPO and INPIT are not responsible for any damages caused by the use of this translation.

1. This document has been translated by computer. So the translation may not reflect the original precisely.
2. **** shows the word which can not be translated.
3. In the drawings, any words are not translated.

CLAIMS

[Claim(s)]

[Claim 1] A node links searching device comprising:

It carries out based on link information showing a name of a node of a link destination and a storing position of a node of a link destination which are included in each node of a hypertext. A search condition setting-out means to set up as the maximum of a hierarchy number which is the number of nodes which exist in from a node used as a starting point of search of a limited condition of a search range at the time of performing search to arbitrary nodes of a link destination from a node one by one before a node of a search place.

A file collecting means which performs repeatedly reading the contents of the node of a link destination which said link information shows from a server which is accumulating it sequentially from a node used as a starting point of said search. A node links information storage means which memorizes information showing matching with a node of a link destination which link information which that node includes whenever it reads the contents of one node by this file collecting means, and this show, and a node read this time.

A hierarchy number calculating means which asks for a hierarchy number of a node of a link destination which link information included in the node shows whenever it reads the contents of one node by said file collecting means, A search range limit means which stops reading by said file collecting means of the contents of the node linked after a node read this time when a hierarchy number for which it asked by this hierarchy number calculating means was larger than the maximum of a hierarchy number set up by said search condition setting-out means.

[Claim 2] A node links searching device comprising:

It carries out based on link information showing a name of a node of a link destination and a storing position of a node of a link destination which are included in each node of a hypertext. A search condition setting-out means to set up as the maximum of a hierarchy number which is the number of links which exist in from a node used as a starting point of search of a limited condition of a search range at the time of performing search to arbitrary nodes of a link destination from a node one by one before a node of a search place.

A file collecting means which performs repeatedly reading the contents of the node of a link destination which said link information shows from a server which is accumulating it sequentially from a node used as a starting point of said search. A node links information storage means which memorizes information showing matching with a node of a link destination which link information which that node includes

whenever it reads the contents of one node by this file collecting means, and this show, and a node read this time.

A hierarchy number calculating means which asks for a hierarchy number of a link in a node of a link destination which link information included in the node expresses whenever it reads the contents of one node by said file collecting means, A search range limit means which stops reading by said file collecting means of the contents of the node linked after a node read this time when a hierarchy number for which it asked by this hierarchy number calculating means was larger than the maximum of a hierarchy number set up by said search condition setting-out means.

[Claim 3]The node links searching device according to claim 1 or 2, wherein a node which is the target of search is distributed and stored in two or more servers connected to a network.

[Claim 4]The node links searching device comprising according to claim 1 or 2:
The same node discriminating means that distinguishes whether a node of a link destination which link information included in a node read this time shows of said file collecting means is the same as that of an already read node.

A multiplex reading stop means to stop reading of that node for the second time when it is distinguished that it is the same as that of a node already read by this same node discriminating means.

[Translation done.]

* NOTICES *

JPO and INPIT are not responsible for any damages caused by the use of this translation.

1. This document has been translated by computer. So the translation may not reflect the original precisely.
2. **** shows the word which can not be translated.
3. In the drawings, any words are not translated.

DESCRIPTION OF DRAWINGS

[Brief Description of the Drawings]

[Drawing 1] It is a block diagram showing the outline of the composition of the node links searching device in one example of this invention.

[Drawing 2] It is an explanatory view showing an example of the contents of registration of a node table.

[Drawing 3] It is an explanatory view showing an example of the contents of registration of a link table.

[Drawing 4] It is a flow chart showing the flow of the operation which a node links searching device performs.

[Drawing 5] It is a flow chart showing the flow of the processing at the time of judging whether it is a node which exists in a search range.

[Drawing 6] It is a flow chart showing the flow of other examples of processing which judge whether it is a node which exists in a search range.

[Drawing 7] It is a flow chart showing the flow of the table update process shown in drawing 4.

[Drawing 8] It is an explanatory view showing an example of the contents of registration of the node table used with the node links searching device of a modification.

[Drawing 9] It is an explanatory view showing an example of the contents of registration of the link table used with the node links searching device of a modification.

[Drawing 10] It is a flow chart showing the flow of the operation which the node links searching device in a modification performs.

[Drawing 11] It is a flow chart showing the flow of the processing at the time of judging whether a hyperlink exists in a search range.

[Drawing 12] It is a flow chart showing the flow of other examples of processing which judge whether a hyperlink exists in a search range.

[Drawing 13] It is a flow chart showing the flow of the table update process shown in drawing 10.

[Drawing 14] It is an explanatory view showing an example of the hyperlink which connects between a node and these nodes.

[Drawing 15] It is an explanatory view showing an example of the relation of the documents of a node.

[Drawing 16] It is a block diagram showing the outline of the composition of the node links searching device currently used from the former.

[Description of Notations]

11 Node links searching device

12 Network
13 Server machine
21 File collecting part
22 Hyperlink extraction part
23 Node links information database
24 Search range control section
25 System control part
26 Hypermedia description language analyzing parts
27 Hierarchy number calculation part
28 Host name comparing element
29 Hierarchy number comparing element
31 and 71 Node table
51 and 81 Link table

[Translation done.]